Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Sound and Vibration 296 (2006) 461–493
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

The time domain response of some systems for sound reproduction ☆

P.A. Nelson*, J.F.W. Rose

Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Received 6 August 2004; received in revised form 8 December 2005; accepted 9 December 2005
Available online 9 June 2006

Abstract

The perception in a listener of the existence of a ''virtual'' source of sound at a prescribed spatial position can be produced by ensuring that the acoustic signals at the listener's ears faithfully replicate those that would be produced by a ''real'' source at the same position. When loudspeakers are used to transmit the signals, it is necessary to pass the signals intended for presentation at the listener's ears through a matrix of filters that provide the inverse of the matrix of transfer functions that relates the loudspeaker input signals to the listener's ear signals. The characteristics of such filter matrices are profoundly influenced by the conditioning of the matrix to be inverted. This filter design problem is reviewed here by representing the loudspeakers as simple point monopole sources the head of the listener as a rigid sphere. The case of a virtual acoustic imaging system that uses two loudspeakers in order to reproduce the signals at the two ears is first described in some detail and previous work is reviewed. It is confirmed that the time domain response of the reproduced field is of long duration at frequencies where the inversion problem is ill-conditioned. The influence of the presence of the listener's head on this time domain behaviour is also evaluated. The principle is then extended to four input–four output reproduction systems and the computational model is used to explain some previous experimental observations. Finally, the conditioning of five input–four output systems is also examined and shown to have some potentially desirable characteristics.

1. Introduction

It is possible, by using a pair of loudspeakers, to produce fluctuating sound pressures at the ears of a listener that replicate accurately a pair of prescribed sound pressure time histories. The latter might be those that would be produced by a particular source of sound located at a specified spatial position relative to the listener. This approach is capable of generating the convincing illusion in the listener of the existence of a virtual source of sound at the specified spatial position. Unlike conventional stereophony [1], the position of the virtual source is not primarily restricted to the range of angular positions in the horizontal plane that falls

☆This paper is the text of the 2002 Rayleigh Medal Lecture presented by the first author to the Institute of Acoustics at *Reproduced Sound 19*, held in Oxford during November 2003. The permission of the Institute of Acoustics to publish this work is gratefully acknowledged.

*Corresponding author. Fax: +44 238 593190.
E-mail address: pan@isvr.soton.ac.uk (P.A. Nelson).

0022-460X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2005.12.051

between the angular positions of the loudspeakers. This approach, based on "cross-talk cancellation", is generally attributed to Atal and Schroeder [2], although Bauer [3] had previously investigated a similar procedure for the reproduction of dummy head recordings. The technique has been further developed by a number of other authors [4–17] and requires the design of a matrix of filters that operates on a pair of binaurally recorded signals (or a pair of binaurally synthesised signals) in order to derive the inputs to the two loudspeakers. This matrix of "cross-talk cancellation filters" effectively inverts the matrix of transfer functions relating the loudspeaker input signals to the listener's ears signals, thus ensuring that the binaurally recorded signals are faithfully replicated at the ears of the listener.

A noteworthy finding [15–17] was that the illusion in the listener was especially convincing when cross-talk cancellation was applied whilst using two loudspeakers that subtend a relatively narrow angle (typically ten degrees) at the position of the listener. This sound reproduction system (dubbed the "Stereo Dipole") was found to have highly desirable properties, especially with regard to the robustness of the performance of the system with respect to the movement of the head of the listener [18]. It was subsequently pointed out by Ward and Elko [19] that the transfer function matrix to be inverted became ill-conditioned when the path-length difference between one of the loudspeakers and the two ears of the listener became equal to one half of the acoustic wavelength. The narrow angular range of the Stereo Dipole thus ensured a well-conditioned inversion problem over a particularly useful range of frequencies.

This concept was extended by Bauck [20] and by Takeuchi et al. [21–24] the latter introducing the concept of the "Optimal Source Distribution" by demonstrating that the inversion problem could be made to be well-conditioned over a very wide range of frequencies by ensuring that the angular span of the loudspeakers was made to vary (preferably continuously) with frequency. A recent analytical investigation [25] of the two source–two field point inversion problem showed clearly how the time domain response of the inverse filters was highly undesirable at the ill-conditioned frequencies, resulting in a sound field with a long duration in the time domain and a complex wave field which would clearly give a deterioration in the cross-talk cancellation performance for small movements of the listeners head. The equivalent analysis in the frequency domain also demonstrated that the spatial extent of cross-talk cancellation was dramatically curtailed at the ill-conditioned frequencies. The analysis presented previously [25] will be extended here by using a model of scattering of sound by the head of a listener based on Lord Rayleigh's analysis of sound interacting with a rigid sphere [26].

Lord Rayleigh used his analysis of spherical scattering at a single frequency in the development of his "Duplex Theory" [27] of sound localisation by the human auditory system. He concluded that the influence of the head at low frequencies (where the wavelength of sound is much larger than the "diameter" of the head) does little to modify the relative amplitude of the sound at the two ears and therefore the available cue for localisation must be the inter-aural time difference (ITD) between the signals arriving at the two ears. By contrast, at higher frequencies (where the acoustic wavelength is comparable to, or much less than, the head diameter), Lord Rayleigh concluded that the inter-aural level difference (ILD) must be the dominant cue for localisation. There is an inherent presumption in this paper that Lord Rayleigh's Duplex theory is incomplete and that ITDs are indeed significant at high frequencies, at least in so far as the time differences between the "envelopes" of high-frequency carrier signals are detectable by the auditory system. There is considerable evidence that this is indeed the case as discussed in a recent review by Hafter and Trahiotis [28]. These authors cite a number of studies that have demonstrated the detectability of ITDs in the envelope of a modulated tone "even when all of the resulting frequencies are above the usual dividing line of the Duplex theory". They also refer to studies showing that ITDs can be detected in bands of high-frequency noise and single and repeated high-frequency clicks, although they do point out that "caution must be applied when designing such experiments because it has been shown that listeners are able to utilise ITDs in the highly attenuated low-frequency skirts of nominally high-frequency signals". Further cautionary comments are made in another review presented by Grantham [29] who cites in particular the work of Wightman and Kistler [30] who concluded that, as far as horizontal plane localisation is concerned, stimuli with low-frequency content are localised primarily on ITDs and stimuli without low-frequency content are localised based primarily on ILDs (including pinna cues). It is interesting to note that this issue does still not appear to have been definitively resolved in the psychoacoustical literature.

2. Scattering of sound by a rigid sphere

Here we make use of the classical solution for the scattering of sound from a rigid sphere in order to provide a reasonable first approximation to the Head Related Transfer Function of the listener. The sound field of a point monopole source having a complex volume velocity q is given by

$$p = \frac{\mathrm{j}\omega\rho_0 q \mathrm{e}^{-\mathrm{j}kr}}{4\pi r}, \tag{1}$$

where p is the complex acoustic pressure, r is the radial distance of the field point from the source, $k = \omega/c_0$ is the wavenumber and ρ_0 and c_0 are, respectively, the density and sound speed of the medium. Denoting $V = \mathrm{j}\omega q$ as the source volume acceleration enables the definition of the frequency response function for free-field radiation given by

$$C_{\mathrm{ff}}(\mathrm{j}\omega) = \frac{p}{V} = \frac{\rho_0 \mathrm{e}^{-\mathrm{j}kr}}{4\pi r}. \tag{2}$$

It is now assumed that this field is radiated by a point source situated relative to a rigid sphere as depicted in Fig. 1. The method of calculating the scattered field is described, for example, by Kirkeby et al. [17]. The expression for the pressure produced by the free-field monopole can be expanded in terms of an infinite series by using well-known series expansions [31] for $\cos(kr)/kr$ and $\sin(kr)/kr$. The expression for the free-field frequency response function relating the acoustic pressure to the volume acceleration of a point monopole source can be written as

$$C_{\mathrm{ff}}(\mathrm{j}\omega) = -\frac{\mathrm{j}\rho_0 k}{4\pi}\sum_{m=0}^{\infty}(2m+1)j_m(kr)\left[j_m(k\rho) - \mathrm{j}n_m(k\rho)\right]P_m(\cos\phi), \tag{3}$$

where the distance ρ and the angle ϕ are defined in Fig. 1, j_m and n_m are, respectively, the mth-order spherical Bessel and Neumann functions and the functions P_m denote the Legendre polynomials of mth-order. The frequency response function relating the scattered field pressure to the volume acceleration of the point monopole can also be expressed in series form by

$$C_s(\mathrm{j}\omega) = \frac{\rho_0 k}{4\pi}\sum_{m=0}^{\infty} b_m\left[j_m(ka) - \mathrm{j}n_m(ka)\right]P_m(\cos\phi), \tag{4}$$

where the coefficients b_m have to be determined, a denotes the radius of the sphere and only outward going waves are assumed. Application of the condition of zero normal pressure gradient on the surface of the sphere shows that the coefficients b_m are given by [17]

$$b_m = \mathrm{j}(2m+1)j'_m(ka)\frac{j_m(kr) - \mathrm{j}n_m(kr)}{j'_m(ka) - \mathrm{j}n'_m(ka)}, \tag{5}$$

where the prime denotes differentiation with respect to the argument of the function.

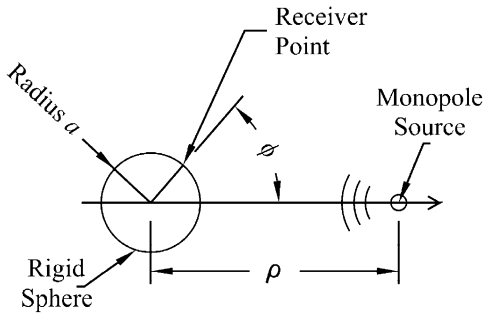

Fig. 1. The coordinate system used to evaluate the scattering of sound by a rigid sphere.

The frequency response function associated with the total field may readily be evaluated by using the MATLAB package, for example, to evaluate a number of terms in the series given by Eq. (4). Rose [32] describes how the spherical Bessel functions may be related to the standard (cylindrical) Bessel functions that are available in MATLAB in order to compute the scattered field efficiently. It has been found that the infinite series given by Eq. (4) converges satisfactorily after about $m = [\text{round}(ka) + 10]$ terms. Some examples of the

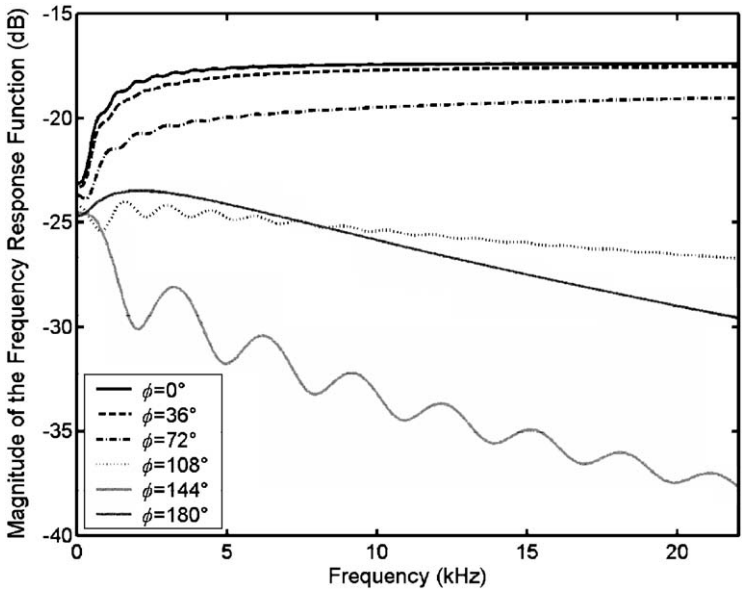

Fig. 2. The frequency response function relating the pressure at a number of angular positions on the surface of a rigid sphere of radius 9 cm to the volume acceleration of a point monopole source at a distance of 1.5 m from the centre of the sphere.

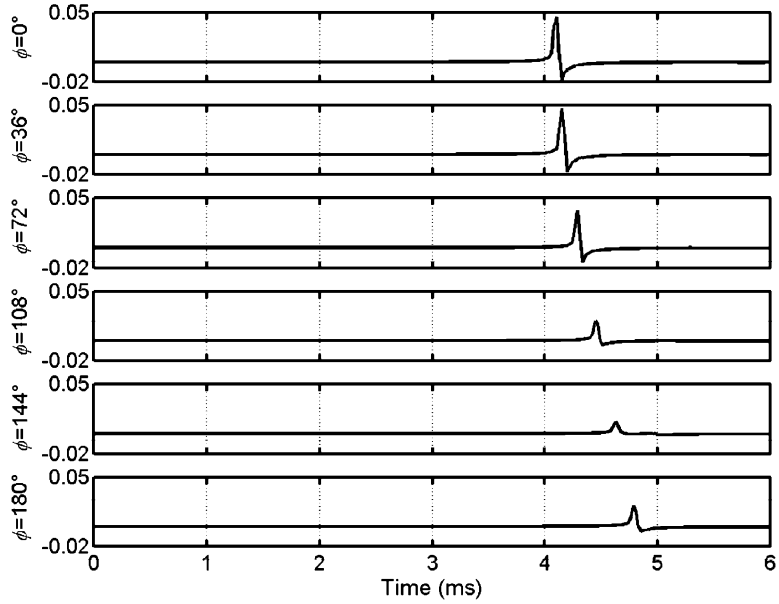

Fig. 3. The impulse responses associated with the frequency response functions plotted in Fig. 2. The frequency response was computed at 1024 discrete frequencies from 1 Hz to 44.1 kHz before transformation into the discrete time domain at an effective sampling frequency of 44.1 kHz.

total frequency response function computed from the sum $C_t(j\omega)$ of $C_s(j\omega)$ and $C_{ff}(j\omega)$ are shown in Fig. 2 for a range of angular positions on the surface of the sphere relative to the position of the monopole source.

The results of this computation may be transformed into equivalent discrete time impulse responses by first windowing this continuous function in the frequency domain and then sampling the frequency response function at N points in the range from $\omega = 0$ to ω_s where the latter denotes an equivalent discrete time sampling frequency. With the discrete frequency variable being denoted by k (which should not be confused with the acoustic wavenumber used above), the discrete time impulse response is computed from the inverse discrete Fourier transform given by

$$c(n) = \frac{1}{N}\sum_{k=0}^{N-1} C(k)e^{j(2\pi nk)/N}, \tag{6}$$

where n denotes the discrete time variable. The discrete time impulse responses corresponding to the frequency response functions of Fig. 2 are shown in Fig. 3. A Hanning window was applied to the frequency response functions of Fig. 2 before transformation into the discrete time domain.

3. Matrix inversion for cross-talk cancellation

The standard signal processing block diagram associated with the multi-channel filter design problem is illustrated in Fig. 4. In this case $\mathbf{C}(k)$ is the matrix of transfer functions in the discrete frequency domain that relates the vector of loudspeaker input signals $\mathbf{v}(k)$ to the vector of signals $\mathbf{w}(k)$ produced at the listeners ears. It is therefore assumed that $\mathbf{w}(k) = \mathbf{C}(k)\mathbf{v}(k)$. The matrix $\mathbf{H}_x(k)$ of cross-talk cancellation filters operates on the vector of (recorded or synthesised) binaural signals $\mathbf{u}(k)$ in order to deduce the vector $\mathbf{v}(k)$ of loudspeaker input signals such that $\mathbf{v}(k) = \mathbf{H}_x\mathbf{u}(k)$. The signals $\mathbf{u}(k)$ could thus be recorded (at the ears of a dummy head for example) or synthesised by convolving the signal associated with an intended virtual source with a pair of filters representing the transfer functions from the position of the virtual source to the ears of the listener. The latter could be, for example, derived from a measured database of Head Related Transfer Functions. The basic requirement of the cross-talk cancellation matrix is to ensure that the reproduced signals are simply a delayed version of the binaural signals. That is we wish to ensure that $\mathbf{w}(k)$ is made equal to the desired signals $\mathbf{d}(k)$ at the listeners ears that are equal to $\mathbf{u}(k)e^{-j\omega\Delta}$ where Δ represents the delay. It therefore follows that the requirement is

$$\mathbf{C}(k)\mathbf{H}_x(k) \approx e^{-j\omega\Delta}\mathbf{I}, \tag{7}$$

where $\mathbf{I}$ denotes the identity matrix. Thus at each frequency k the solution for the cross-talk cancellation matrix is in principle given by

$$\mathbf{H}_x(k) \approx \mathbf{C}^{-1}(k)e^{-j\omega\Delta}, \tag{8}$$

It is interesting to note first the analytical form that this matrix takes when it is assumed that the matrix $\mathbf{C}(k)$ is that associated with an arrangement of two loudspeakers placed symmetrically with respect to a listener. Thus, assume that

$$\mathbf{C}(k) = \begin{bmatrix} C_D(k) & C_C(k) \\ C_C(k) & C_D(k) \end{bmatrix}, \tag{9}$$

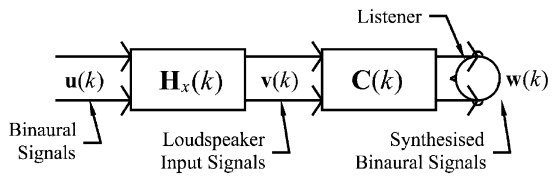

Fig. 4. The multi-channel signal processing block diagram associated with the cross-talk cancellation problem.

where $C_D(k)$ and $C_C(k)$ are, respectively, the frequency responses of the "direct" and "cross-talk" transmission paths from one of the loudspeakers to the nearest and furthest ear of the listener. The inverse of this matrix is given by

$$\mathbf{C}^{-1}(k) = \frac{1}{C_D^2(k) - C_C^2(k)} \begin{bmatrix} C_D(k) & -C_C(k) \\ -C_C(k) & C_D(k) \end{bmatrix}. \tag{10}$$

Writing the ratio of the cross-talk to the direct path as $C_C(k)/C_D(k) = R(k)$ shows that this expression can be written as

$$\mathbf{C}^{-1}(k) = \frac{1}{C_D(k)\{(1 - R(k))(1 + R(k))\}} \begin{bmatrix} 1 & -R(k) \\ -R(k) & 1 \end{bmatrix}. \tag{11}$$

Now assume as a first approximation that the two transmission paths are governed only by the propagation delay and amplitude reduction associated with the spherical spreading of sound from a point monopole source as defined by Eq. (1) above. Thus, for example, if

$$C_D(k) = \frac{\rho_0 e^{-j\omega r_1/c_0}}{4\pi r_1}, \quad C_C(k) = \frac{\rho_0 e^{-j\omega r_2/c_0}}{4\pi r_2}, \tag{12a,b}$$

then we may write $R(k) = g e^{-j\omega\tau}$ where $g = r_2/r_1$ is the ratio of the two path lengths and $\tau = (r_2 - r_1)/c_0$ is the difference between the acoustic travel times from one of the loudspeakers to the furthest and nearest ears of the listener.

It can therefore be deduced that in this simple case, it is in principle possible to build a realisable matrix of filters in order to accomplish the cross-talk cancellation. The only component that cannot be realised in Eq. (11) is the term $1/C_D(k)$. This also clearly illustrates the necessity of the "modelling delay" introduced into the numerator of the solution for $\mathbf{H}_x(k)$ by the term $e^{-j\omega\Delta}$ since, provided the delay Δ exceeds the delay r_1/c_0 then there is no requirement to implement a "time advance". The terms $R(k)$ appearing in the numerator of equation (11) (i.e. in the adjoint matrix) are clearly realisable since they constitute pure delays and the terms $1/(1 + R(k))$ and $1/(1 - R(k))$ could also be realised in discrete time as recursive filters (since again the elements associated with $R(k)$ are a pure delay).

Note, however, a potential difficulty with the implementation of the filters defined in Eq. (11). In particular, note the modulus squared of the filters appearing in the denominator. These can be written as

$$\left|1 - R(k)\right|^2 = \left(1 + g^2 - 2\cos\omega\tau\right), \quad \left|1 + R(k)\right|^2 = \left(1 + g^2 + 2\cos\omega\tau\right). \tag{13}$$

Since the ratio g will generally be close to unity, these terms will become small as the frequency ω tends to zero and at frequencies where $\cos\omega\tau = 1$ or -1, respectively. This occurs when $\omega\tau = n\pi$ where n is an integer. The response of the filters specified by Eq. (11) therefore becomes large at these frequencies, the frequency at which $\omega\tau = \pi$ or $f = 1/2\tau$ being the "ringing frequency" identified by Kirkeby et al. [16] as being associated with an undesirable response in the time domain.

4. Ill-conditioning of the inversion problem

As described in detail in the analytical investigation of the two source–two field point inversion problem [25], the ringing frequency is associated with ill-conditioning of the frequency response function matrix to be inverted, a complex sound field in the region of the listener's head in the time domain and an associated reduction in the spatial extent of cross-talk cancellation. The condition number of the matrix $\mathbf{C}(k)$ to be inverted is defined in terms of the singular value decomposition (SVD) of the matrix that can be written in the form

$$\mathbf{C}(k) = \mathbf{U}(k)\mathbf{\Sigma}(k)\mathbf{V}^{\mathrm{H}}(k), \tag{14}$$

where $\mathbf{\Sigma}(k)$ is the diagonal matrix of singular values, $\mathbf{U}(k)$ and $\mathbf{V}(k)$ are the unitary matrices of left and right singular vectors, respectively, and the superscript H denotes the Hermitian (complex conjugate) transpose. The condition number $\kappa(\mathbf{C})$ of the matrix $\mathbf{C}(k)$ is given by the ratio of the maximum to minimum singular

values that comprise the elements of the diagonal matrix $\mathbf{\Sigma}$. The significance of the condition number in dealing with matrix inversion problems is of course well-known [33]. Assuming one determines the loudspeaker input signals $\mathbf{v}(k)$ from the solution for the cross-talk cancellation matrix given by $\mathbf{H}_x(k) \approx \mathbf{C}^{-1}(k)\mathrm{e}^{-\mathrm{j}\omega\Delta}$, it follows that

$$\mathbf{v}(k) = \mathbf{C}^{-1}(k)\mathbf{d}(k) = \mathbf{C}^{-1}(k)\mathbf{u}(k)\mathrm{e}^{-\mathrm{j}\omega\Delta}. \tag{15}$$

It can then be shown that [33] the errors $\delta\mathbf{v}(k)$ in the solution for $\mathbf{v}(k)$ are related to the errors $\delta\mathbf{C}(k)$ in the specification of the matrix $\mathbf{C}(k)$ and the errors $\delta\mathbf{d}(k)$ in the specification of the desired signals $\mathbf{d}(k)$ at the listeners ears by the inequality

$$\frac{\|\delta\mathbf{v}(k)\|}{\|\mathbf{v}(k)\|} \leqslant \kappa(\mathbf{C})\left[\frac{\|\delta\mathbf{C}(k)\|}{\|\mathbf{C}(k)\|} + \frac{\|\delta\mathbf{d}(k)\|}{\|\mathbf{d}(k)\|}\right]. \tag{16}$$

In this expression, the symbol $\|\ \|$ denotes the 2-norm (which is the sum of the squared elements of a vector or the square root of the largest singular value of a matrix). The errors in the solution for the loudspeaker input voltages resulting from other inaccuracies (in recording or synthesising the binaural signals used to specify the vector $\mathbf{d}(k)$ or in measuring the matrix $\mathbf{C}(k)$) can be amplified by the condition number of the matrix to be

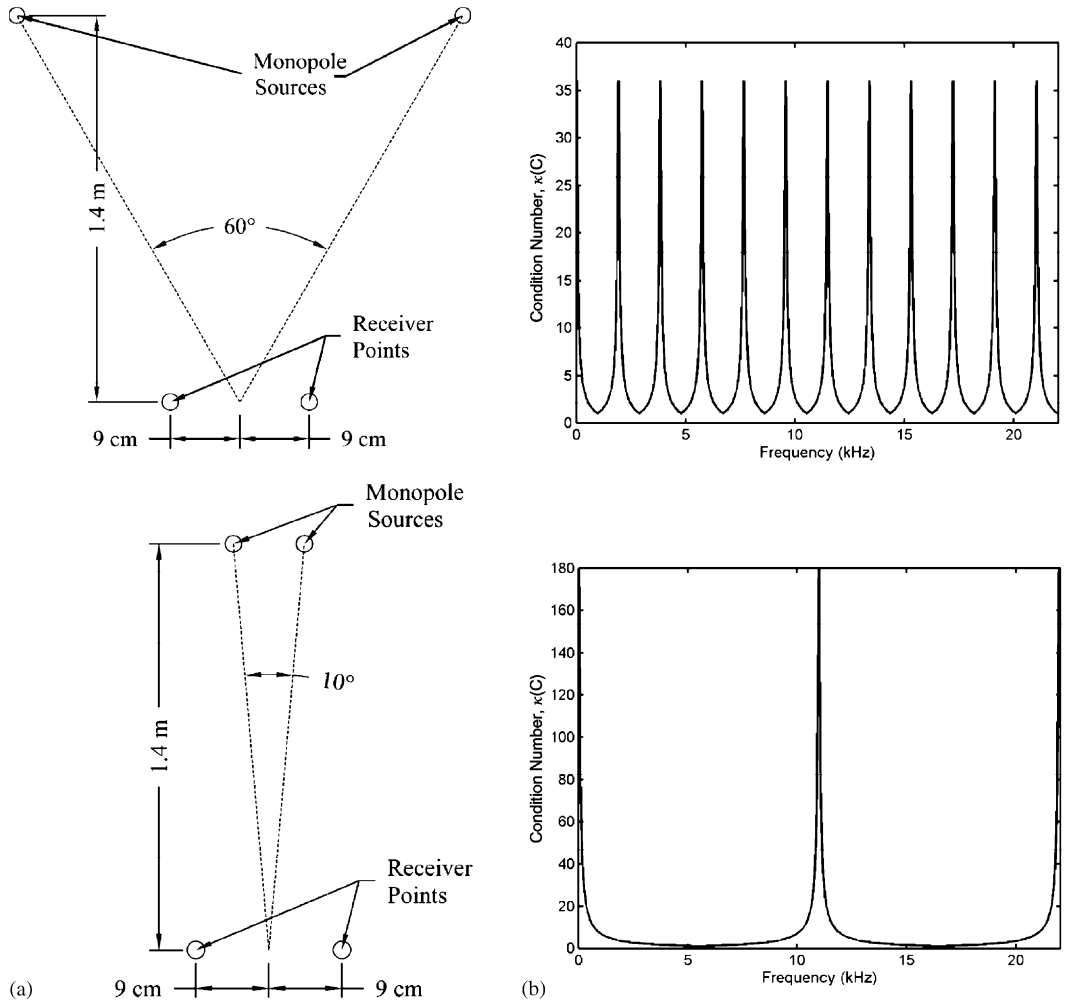

Fig. 5. The condition number of the matrix $\mathbf{C}(k)$ is shown in (b) for the free-field geometrical arrangements of sources and listener shown in (a).

inverted. Large condition numbers can therefore lead to large errors in the solution. As emphasised above, in the case of the free-field two source–two field point problem large condition numbers are also associated with "ringing" in the sound field and deterioration in the spatial extent of cross-talk cancellation.

In the case of two loudspeakers placed symmetrically relative to the listener, and again assuming free-field transfer functions as defined in Eq. (1) above, it can be shown [24,25] that the SVD of the matrix $\mathbf{C}(k)$ results in the following:

$$\mathbf{\Sigma}(k) = C_D(k)\begin{bmatrix} |1 + R(k)| & 0 \\ 0 & |1 - R(k)| \end{bmatrix}, \tag{17}$$

$$\mathbf{U}(k) = \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{\frac{1+R(k)}{1+R^*(k)}} & \sqrt{\frac{1-R(k)}{1-R^*(k)}} \\ \sqrt{\frac{1+R(k)}{1+R^*(k)}} & -\sqrt{\frac{1-R(k)}{1-R^*(k)}} \end{bmatrix}, \tag{18}$$

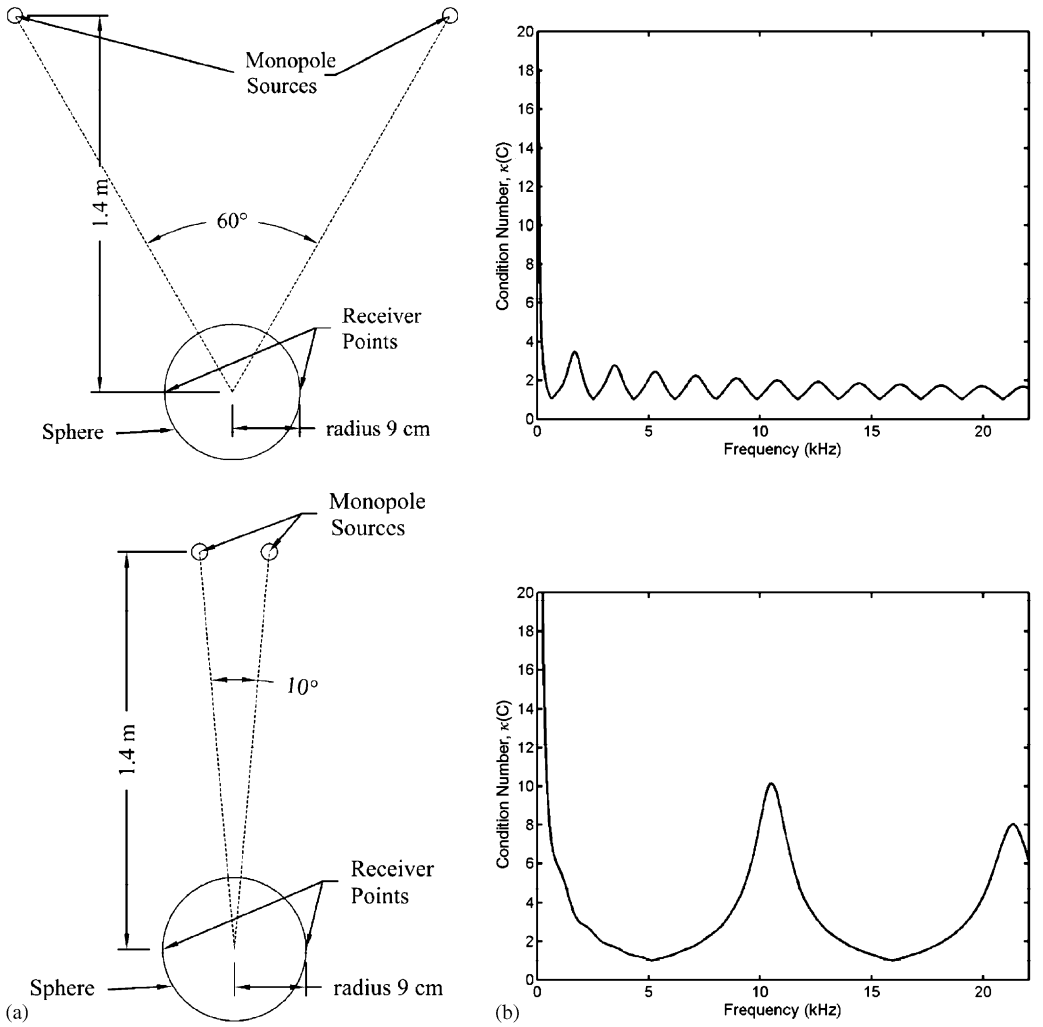

Fig. 6. The condition number of the matrix $\mathbf{C}(k)$ is shown in (b) for the geometrical arrangement of sources and listener shown in (a) where the listeners head is modelled by a rigid sphere.

$$\mathbf{V}(k) = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{19}$$

Fig. 5(b) shows the condition number of $C(k)$ that is given by the ratio of the maximum singular value to the minimum singular value when the two point sources subtend an angle of 60° at the listener in the particular geometrical arrangement shown in Fig. 5(a). Note that the condition number is effectively given by the ratio $|1 + R(k)|/|1 - R(k)|$. As described in Ref. [25], the peaks in the condition number occur at frequencies where the path-length difference $(r_2 - r_1)$ either tends to zero or is equal to integer multiples of one half of the acoustic wavelength. The first maximum in the condition number corresponding to a path-length difference of one half of an acoustic wavelength and defines the "ringing frequency" [16,25]. The effect of narrowing the transducer span to 10° is also shown in Fig. 5(b), which illustrates that the ringing frequency is pushed much higher, but also shows that the conditioning problem at low frequencies becomes more severe.

Finally, note that the inverse of the acoustic transfer function matrix can be written in terms of the SVD by using the unitary properties of the matrices $\mathbf{U}(k)$ and $\mathbf{V}(k)$ such that

$$\mathbf{C}^{-1}(k) = \mathbf{U}(k)\mathbf{\Sigma}^{-1}(k)\mathbf{V}^{\mathrm{H}}(k), \tag{20}$$

where the inverse of the matrix of singular values can be written as

$$\mathbf{\Sigma}^{-1} = \frac{1}{C_D(k)}\begin{bmatrix} \frac{1}{|1+R(k)|} & 0 \\ 0 & \frac{1}{|1-R(k)|} \end{bmatrix}. \tag{21}$$

This again illustrates the tendency of the filters to "blow up" in magnitude in the low-frequency limit and at the frequencies defined above by Eq. (13).

The presence of the head of the listener will of course modify the conditioning of the system. The effect of the head of the listener can be estimated by using the model of scattering by a rigid sphere described above. Here, we assume that the ears of the listener are located at angular positions as illustrated in Fig. 6(a). When the matrix $\mathbf{C}(k)$ is calculated using this model, the resulting condition number as a function of frequency is plotted in Fig. 6(b) for loudspeaker spans of both 60° and 10°. This demonstrates that the peak at the ringing frequency is suppressed for a span of 60°, but at a span of 10° the ringing frequency becomes more pronounced and the condition number becomes higher at low frequencies.

5. Regularised inversion

A well-established technique for dealing with ill-conditioned inversion problems is the use of regularisation. The solution for the vector $\mathbf{v}(k)$ is sought that minimises the sum of the squares of the errors $\mathbf{e}(k)$ (equal to $\mathbf{d}(k) - \mathbf{w}(k)$) between the desired signals $\mathbf{d}(k)$ at the listeners ears and the reproduced signals $\mathbf{w}(k)$, plus a term $\mathbf{v}^{\mathrm{H}}(k)\mathbf{v}(k)$ equal to the sum of squared loudspeaker input voltages weighted by a factor β . The solution [34] to this optimisation problem is given by the optimal vector of loudspeaker input signals defined by

$$\mathbf{v}_{\mathrm{opt}}(k) = \left[\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k) + \beta\mathbf{I}\right]^{-1}\mathbf{C}^{\mathrm{H}}(k)\mathbf{d}(k), \tag{22}$$

where β is the regularisation parameter. Since $\mathbf{d}(k)$ is equal to $\mathbf{u}(k)\mathrm{e}^{-\mathrm{j}\omega\Delta}$, the cross-talk cancellation matrix in Eq. (7) can be written in terms of the regularised pseudo-inverse matrix $\left[\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k) + \beta\mathbf{I}\right]^{-1}\mathbf{C}^{\mathrm{H}}(k)$ such that

$$\mathbf{H}_{xR}(k) = \left[\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k) + \beta\mathbf{I}\right]^{-1}\mathbf{C}^{\mathrm{H}}(k)\mathrm{e}^{-\mathrm{j}\omega\Delta}. \tag{23}$$

Substitution of the SVD given by Eq. (14) into this solution and use of the unitary properties of $\mathbf{U}(k)$ and $\mathbf{V}(k)$ shows that

$$\mathbf{H}_{xR}(k) = \mathbf{V}(k)\left[\Sigma^{\mathrm{H}}(k)\Sigma(k) + \beta\mathbf{I}\right]^{-1}\Sigma^{\mathrm{H}}(k)\mathbf{U}^{\mathrm{H}}(k)\mathrm{e}^{-\mathrm{j}\omega\Delta}. \tag{24}$$

Again using for illustrative purposes the free-field two-source two-field point model described above, it follows that

$$\left[\Sigma^{\mathrm{H}}(k)\Sigma(k)+\beta\mathbf{I}\right]^{-1}\Sigma^{\mathrm{H}}(k)=\frac{1}{C_D(k)}\begin{bmatrix}\frac{|1+R(k)|}{|1+R(k)|^2+\beta} & 0\\ 0 & \frac{|1-R(k)|}{|1-R(k)|^2+\beta}\end{bmatrix} \tag{25}$$

and therefore the presence of the regularisation parameter limits the magnitude of the cross-talk cancellation filters at the particular frequencies where the terms $|1+R(k)|^2$ and $|1-R(k)|^2$ become small.

6. The design of cross-talk cancellation filters

It turns out that the regularised solution given by Eq. (23) provides an extremely convenient practical technique for designing the cross-talk cancellation matrix in particular and inverse filters generally [35]. The central problem associated with inverse filter design is that if the transfer function to be inverted ($C(k)$ say) is non-minimum phase (i.e. has zeros outside the unit circle in the complex z-plane), then the inverse filter ($1/C(k)$) will be unstable (since the zeros outside the unit circle become poles of the inverse filter). However, it can also be demonstrated that poles outside the unit circle can be interpreted as contributing a stable but anti-

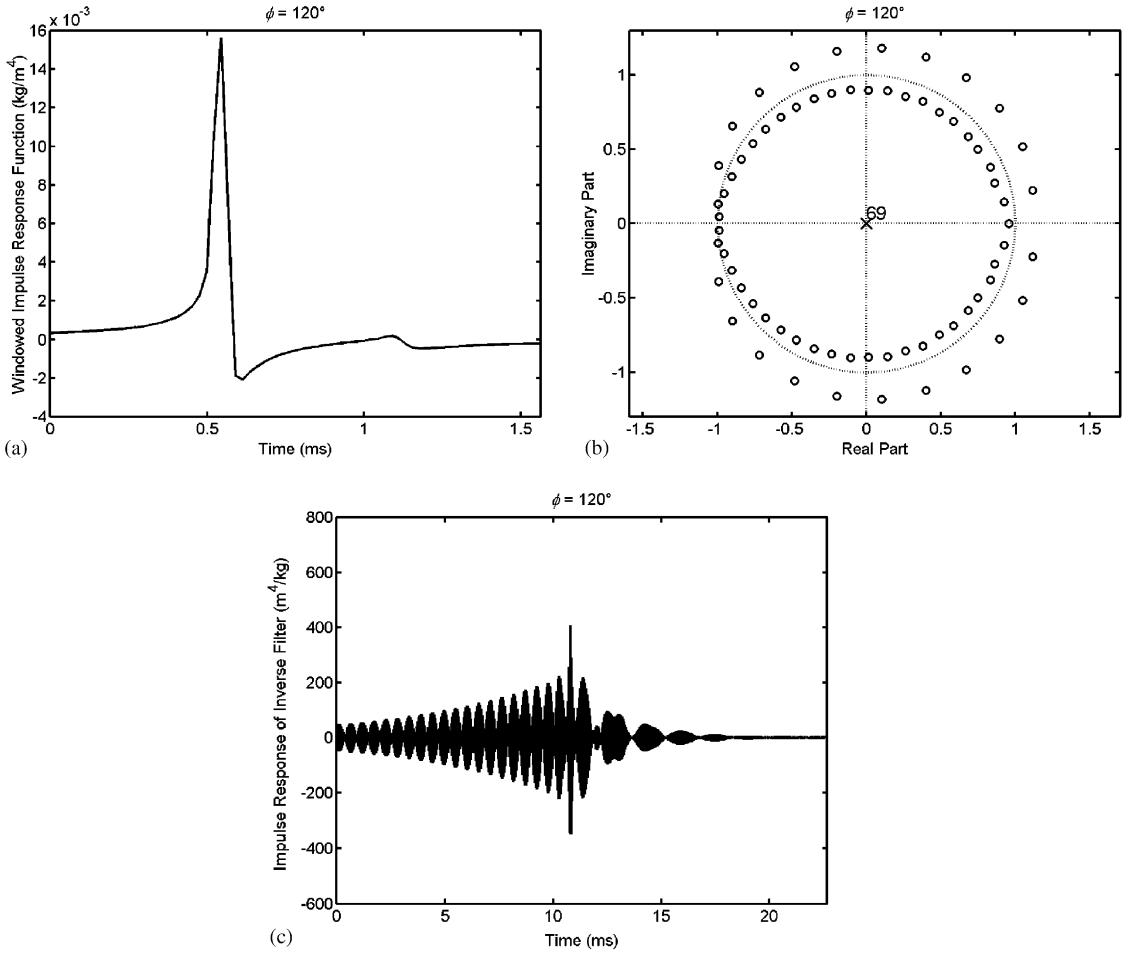

Fig. 7. The (a) impulse response (b) pole-zero map and (c) impulse response of the inverse filter associated with a typical acoustic transfer function. The transfer function shown relates the pressure at the "furthest ear" to the volume acceleration of one of the sources including an angle of 60° in the geometrical arrangement shown in Fig. 6(a).

causal component of the impulse response of the inverse filter. As an example, consider the impulse response $c(n)$ associated with the transfer function $C(k)$ relating the volume acceleration of a point monopole source to the pressure at given angular location on the surface of a rigid sphere. The impulse response $c(n)$ has been computed as described above and the corresponding sequence can be described in terms of the z-transform

$$C(z) = z^{-q}B(z), \quad B(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + \cdots + b_{N-1} z^{-(N-1)}, \tag{26a,b}$$

where q denotes the number of samples delay resulting from the acoustic propagation. The values of the terms b_n are the values of the non-zero terms in the impulse response $c(n)$ and N denotes the number of coefficients used to represent the impulse response (which was chosen here to be 70). The polynomial $B(z)$ can then be factored into a product of the form

$$B(z) = b_0\left(1 - z_1 z^{-1}\right)\left(1 - z_2 z^{-2}\right)\ldots\left(1 - z_N z^{-1}\right), \tag{27}$$

where the terms z_i denote the zeros of the polynomial $B(z)$. Fig. 7(a) shows the impulse response $c(n)$ corresponding to one of the acoustic transmission paths associated with the source receiver geometry of Fig. 6(a). The impulse response illustrated is that associated with the path from one of the sources to the furthest ear with the sources including an angle of 60°. The bulk delay has been removed from the impulse response. The zeros of the associated polynomial $B(z)$ are shown plotted in the z-plane in Fig. (7b). It is clearly evident that there are a number of zeros outside the unit circle defined by $|z| = 1$ and that this representation of the

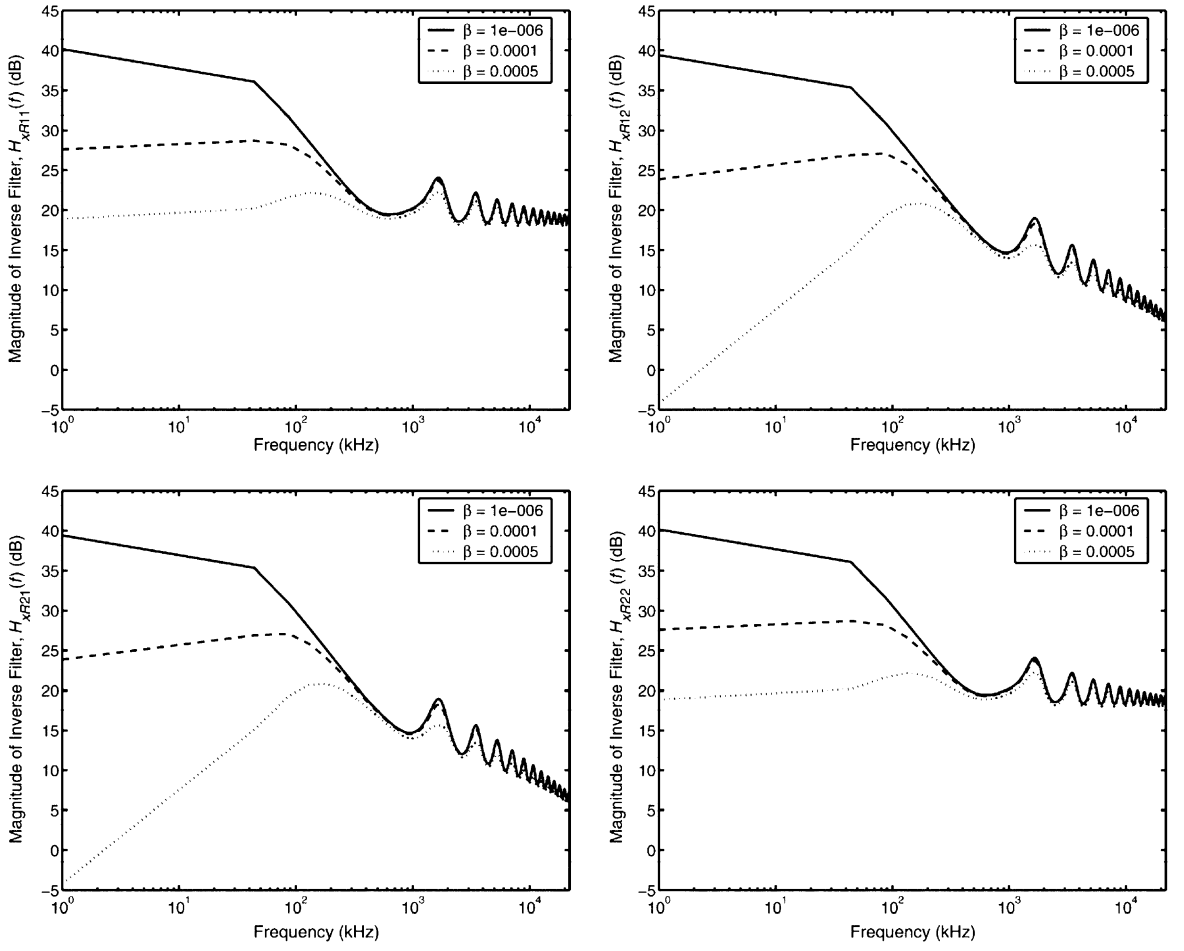

Fig. 8. The frequency response functions of the inverse filters designed for the 60° source angle shown in the geometry of Fig. 6(a). Shown for values of regularisation parameter $\beta = 10^{-6}$, 0.0001 and 0.0005.

transfer function is non-minimum phase. Note also that since the coefficients b_n are real then the zeros z_i must either real or appear in complex conjugate pairs.

The inverse of the transfer function $B(z)$ can be expressed as a partial fraction expansion of the form

$$B^{-1}(z) = \frac{A_1}{1 - z_1 z^{-1}} + \frac{A_2}{1 - z_2 z^{-1}} + \cdots + \frac{A_N}{1 - z_N z^{-1}}, \tag{28}$$

where it has been assumed that the poles of this expression are distinct. It can be argued [36], that when the inverse z-transform is evaluated of each term in this series, the resulting time domain sequence will be determined by the position of the relevant zero z_i relative to the unit circle $|z| = 1$ in the z-plane. Thus, in the case of zeros inside the unit circle, it can be shown [36] that the inverse z-transform yields a causal sequence that decays exponentially in forward time. Conversely, for zeros outside the unit circle, it can be argued that the inverse z-transform yields an anti-causal sequence that decays in backward time. Finally, zeros that lie on the unit circle will result in a sequence of infinite duration in forward or backward time. It should also be noted that the same analysis demonstrates that the closer is the zero to the unit circle, the slower the rate of decay of the impulse response in either forward or backward time. The impulse response of the inverse filter $B^{-1}(z)$ that inverts the impulse response of Fig. 7(a) is illustrated in Fig. 7(c). This was evaluated numerically using the MATLAB package to evaluate the partial fraction expansion. Such a procedure becomes extremely

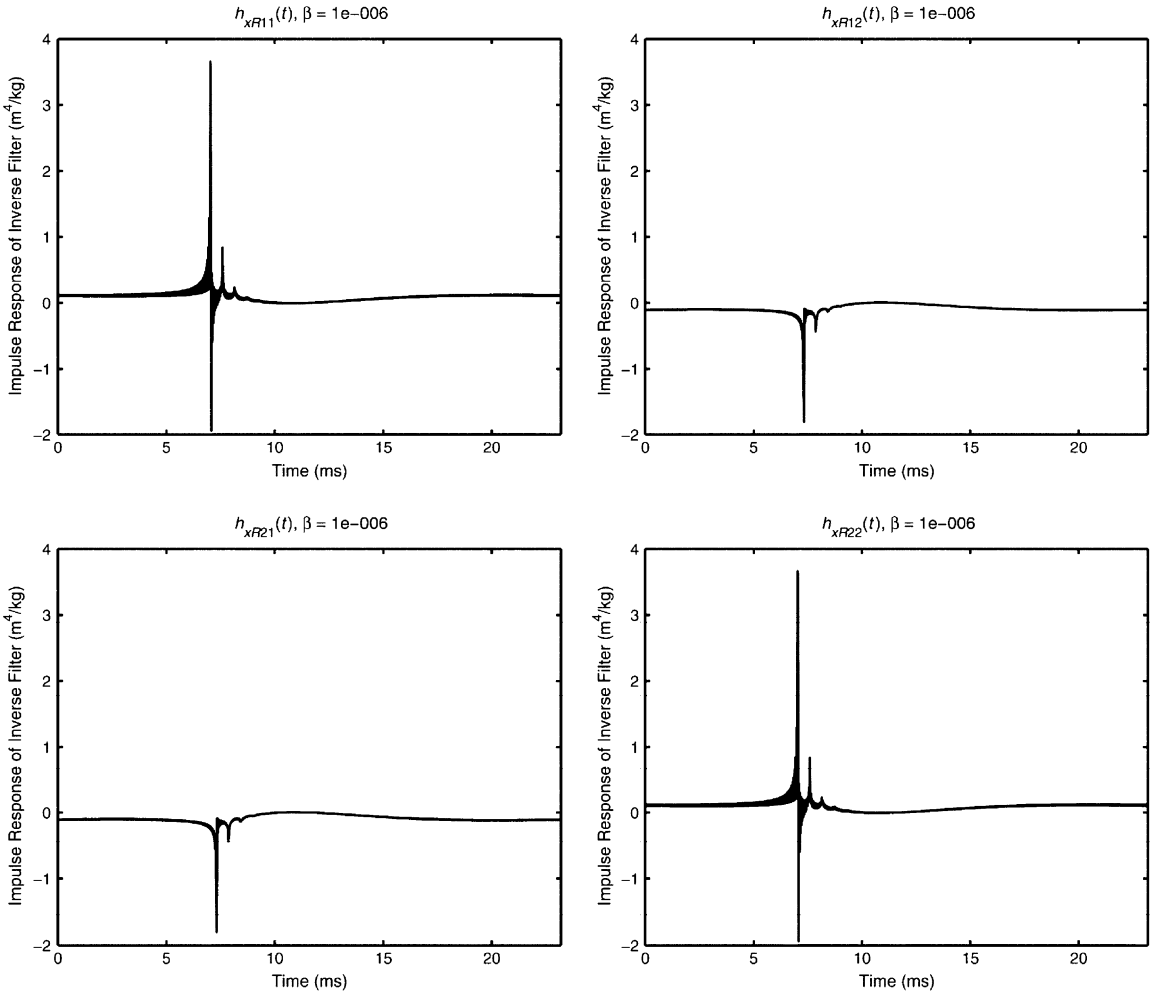

Fig. 9. The impulse responses of the of the inverse filters designed for the 60° source angle shown in the geometry of Fig. 6(a). Shown for values of regularisation parameter $\beta = 10^{-6}$.

difficult for plants of much higher order than the 70 terms used here. However, it was shown by Kirkeby et al. [35] that the problem of inverting multi-channel non-minimum phase systems is dealt with efficiently by using the regularised solution for the cross-talk cancellation matrix provided by Eq. (23). The procedure works entirely in the discrete frequency domain and uses the regularisation parameter β to effectively control the rate of decay of the impulse responses of the inverse filters. The action of the regularisation parameter is to replace each zero with a pair of zeros that are each further away from the unit circle in the z-plane. It is thus ensured that the response is contained within a duration that is sufficiently short compared to the length of the discrete Fourier transform used (whose spectrum of course repeats periodically). The effects of "wrap-around" errors are thus minimised. Figs. 8–13 show the frequency responses and impulse responses of the matrices of inverse filters designed to achieve cross-talk cancellation for both geometries of sources (10°and 60°) shown in Fig. 6. The effect of the variation in regularisation parameter is shown in the frequency responses but only the impulse responses associated with a specific value of β are shown. Also shown is the effectiveness of cross-talk cancellation in each case as β is varied. This is shown with plots of the frequency response functions of the elements of the control performance matrix given by the product $\mathbf{C}(k)\mathbf{H}_{xR}(k)$. If we define this matrix as

$$\mathbf{P}(k) = \mathbf{C}(k)\mathbf{H}_{xR}(k) = \begin{bmatrix} P_{11}(k) & P_{12}(k) \\ P_{21}(k) & P_{22}(k) \end{bmatrix}, \tag{29}$$

then one would expect that perfect cross-talk cancellation would result in unit values of $P_{11}(k)$ and $P_{22}(k)$ and zero values of $P_{12}(k)$ and $P_{21}(k)$. The results in Figs. 8–13 show the elements of this matrix on a logarithmic

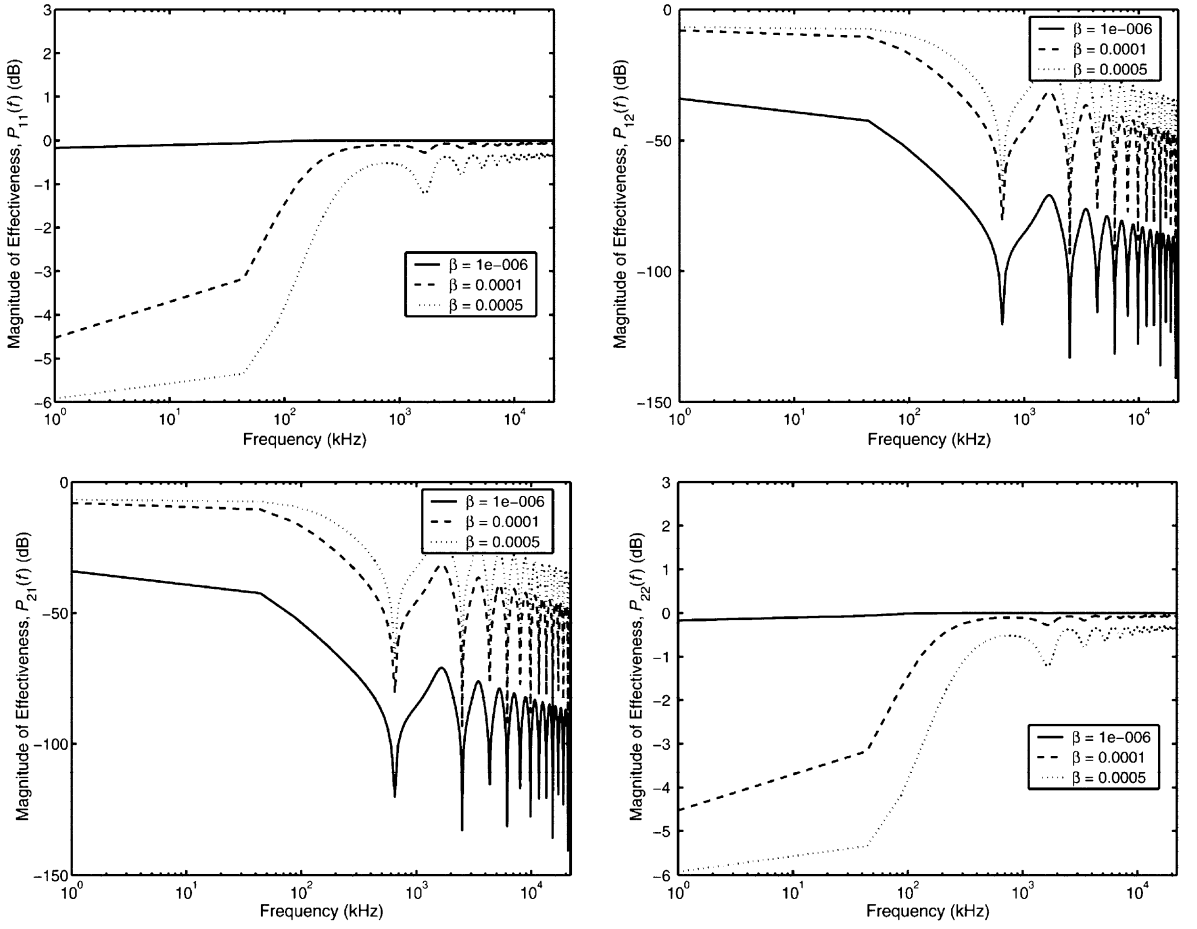

Fig. 10. The effectiveness of cross-talk cancellation of the inverse filters designed for the 60° source angle shown in the geometry of Fig. 6(a). Shown for values of regularisation parameter $\beta = 10^{-6}$, 0.0001 and 0.0005.

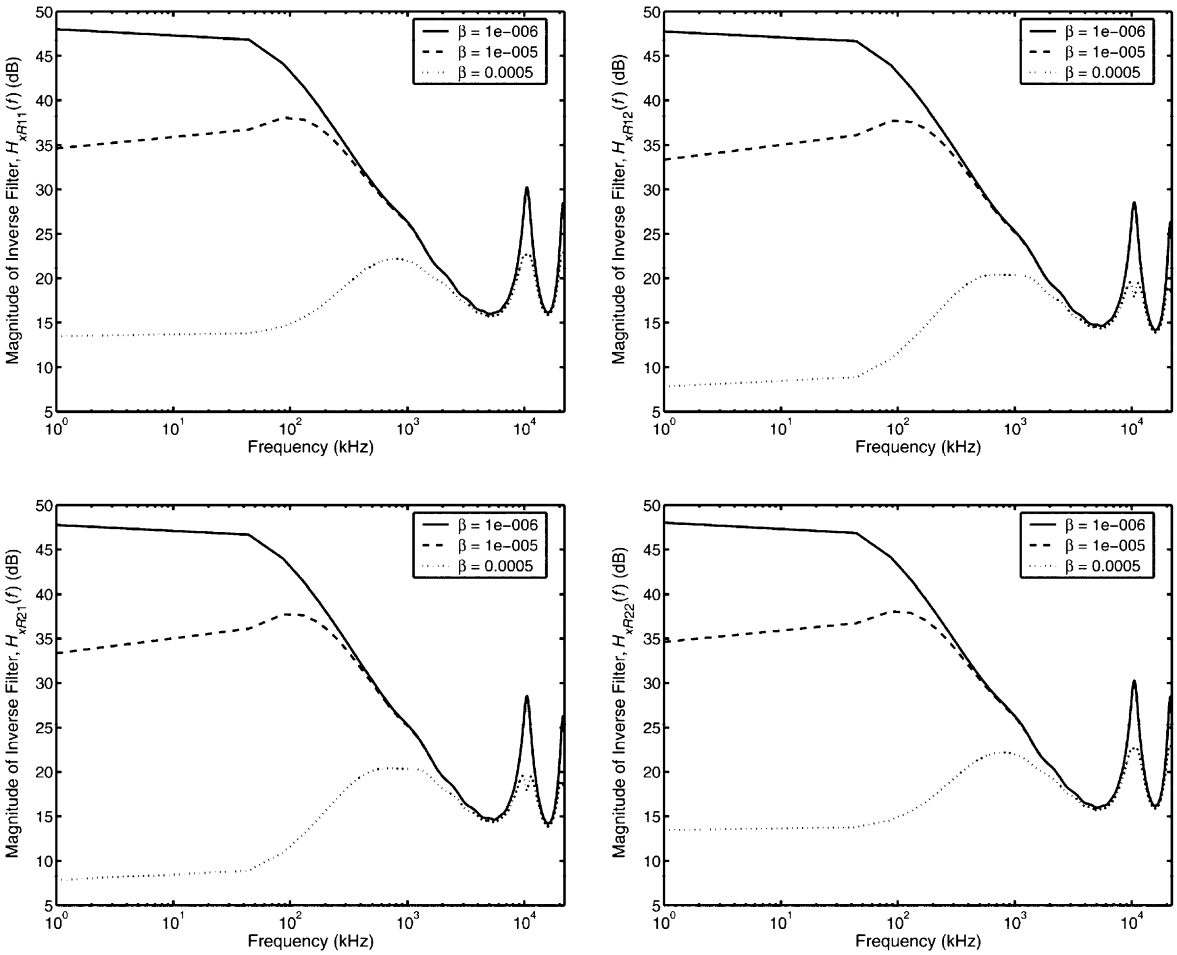

Fig. 11. The frequency response functions of the inverse filters designed for the 10° source angle shown in the geometry of Fig. (6a). Shown for values of regularisation parameter $\beta = 10^{-6}$, 10^{-5} and 0.0005.

scale and demonstrate that the reduction in the gain of the cross-talk cancellation filters that is produced by increasing regularisation results in a deterioration in performance (as one might anticipate). There is obviously a trade off between "control performance" and "control effort" that can be adjusted in a very simple way through choice of the regularisation parameter.

7. The time domain response of two-channel systems

The significance of the ringing frequency to the form of the sound field can be best illustrated by observing the behaviour of the system in the time domain. A previous paper [25] illustrated effectively the time domain response of the free-field two-source two-field point system by computing the source outputs when the signal desired at one field point (ear) was a short duration pulse whilst the desired signal at the other field point (ear) was zero. The pulse chosen to illustrate the system response was the Gaussian signal defined by

$$d(t) = \mathrm{e}^{-\pi(at)^2} \cos \omega_0 t, \tag{30}$$

that has the Fourier transform

$$D(\mathrm{j}\omega) = \frac{1}{2|a|}\left[\mathrm{e}^{-(\omega-\omega_0)^2/4\pi a^2} + \mathrm{e}^{-(\omega+\omega_0)^2/4\pi a^2}\right]. \tag{31}$$

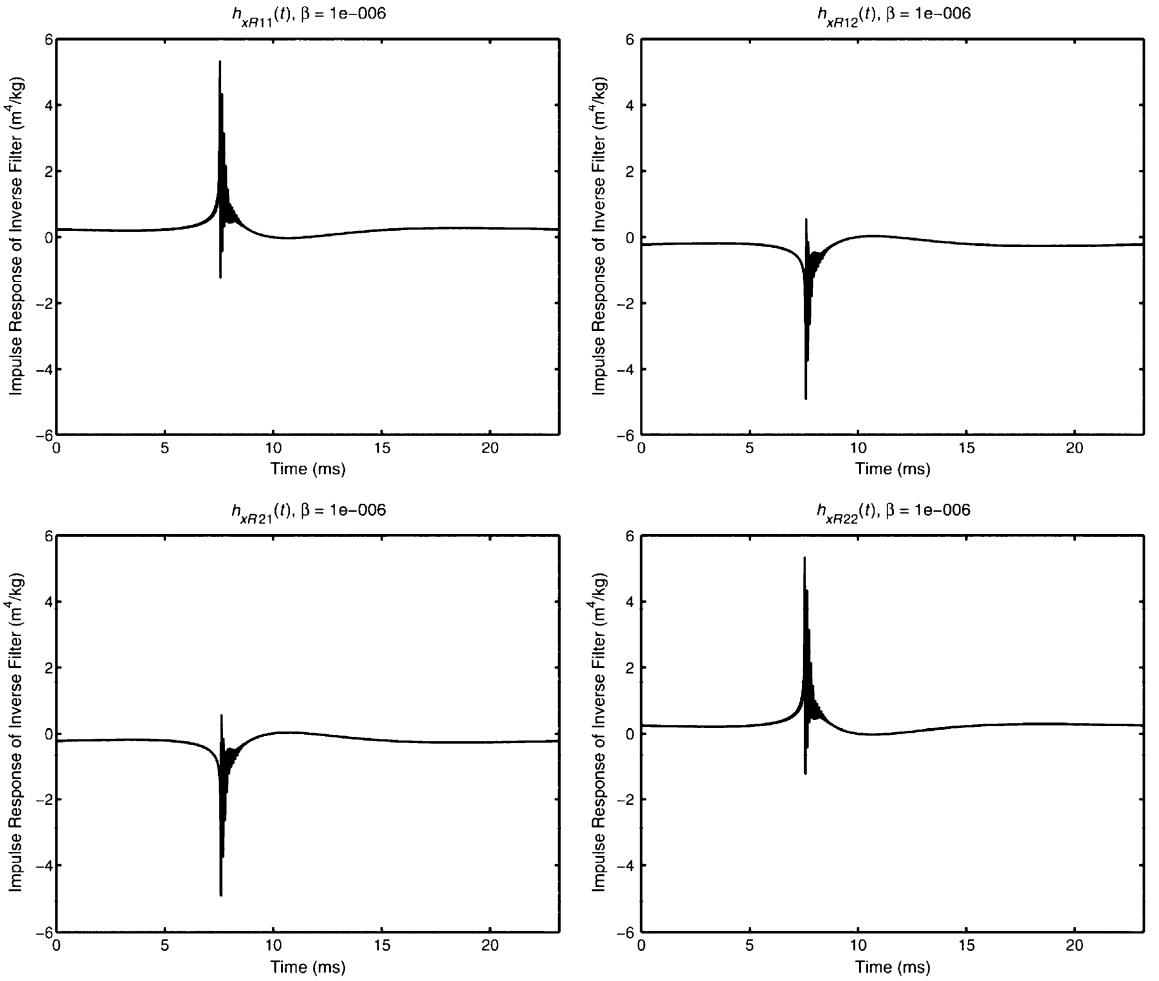

Fig. 12. The impulse responses of the inverse filters designed for the 10° source angle shown in the geometry of Fig. (6a). Shown for values of regularisation parameter $\beta = 10^{-6}$.

The advantage of this waveform is that the centre frequency ω_0 can be chosen to be in a particular frequency range of interest and the spectral width of this pulse can be controlled by the parameter a. The larger the value of a chosen, the narrower is the spectrum of the pulse whilst the longer is the duration of the pulse in the time domain. The pulse also has the minimum product of bandwidth and duration necessary to satisfy the uncertainty principle [37]. In order to examine the behaviour of systems designed to include the effect of listener head scattering, the matrix $\mathbf{C}(k)$ is first computed from the spherical scattering model and the matrix of cross-talk cancellation filters is computed using the discrete Fourier transform as described above. The resulting signals (in discrete time) are then computed by sampling the waveform defined above in order to define the first element of the vector of binaural sequences $\mathbf{u}(n)$, with the second element being set to zero. The discrete Fourier transform of this sequence is then multiplied with the matrix of frequency responses of the cross-talk cancellation filters $\mathbf{H}_{xR}(k)$ in order to deduce the "loudspeaker inputs" $\mathbf{v}(k)$. Multiplication of these by the frequency responses computed from the scattering model described above then yields the net frequency response at each point in the sound field. An inverse discrete Fourier transform is then applied in order to produce a series of "snapshots" of the sound field that results as a function of time.

Some examples of the resulting sound fields are shown in Figs. 14 and 15. These figures represent acoustic compressions as darker areas and rarefactions as lighter. The series of snapshots in Fig. 14 show the result of the attempting to produce the desired Gaussian pulse at the "left ear" of the listener and zero signal at the

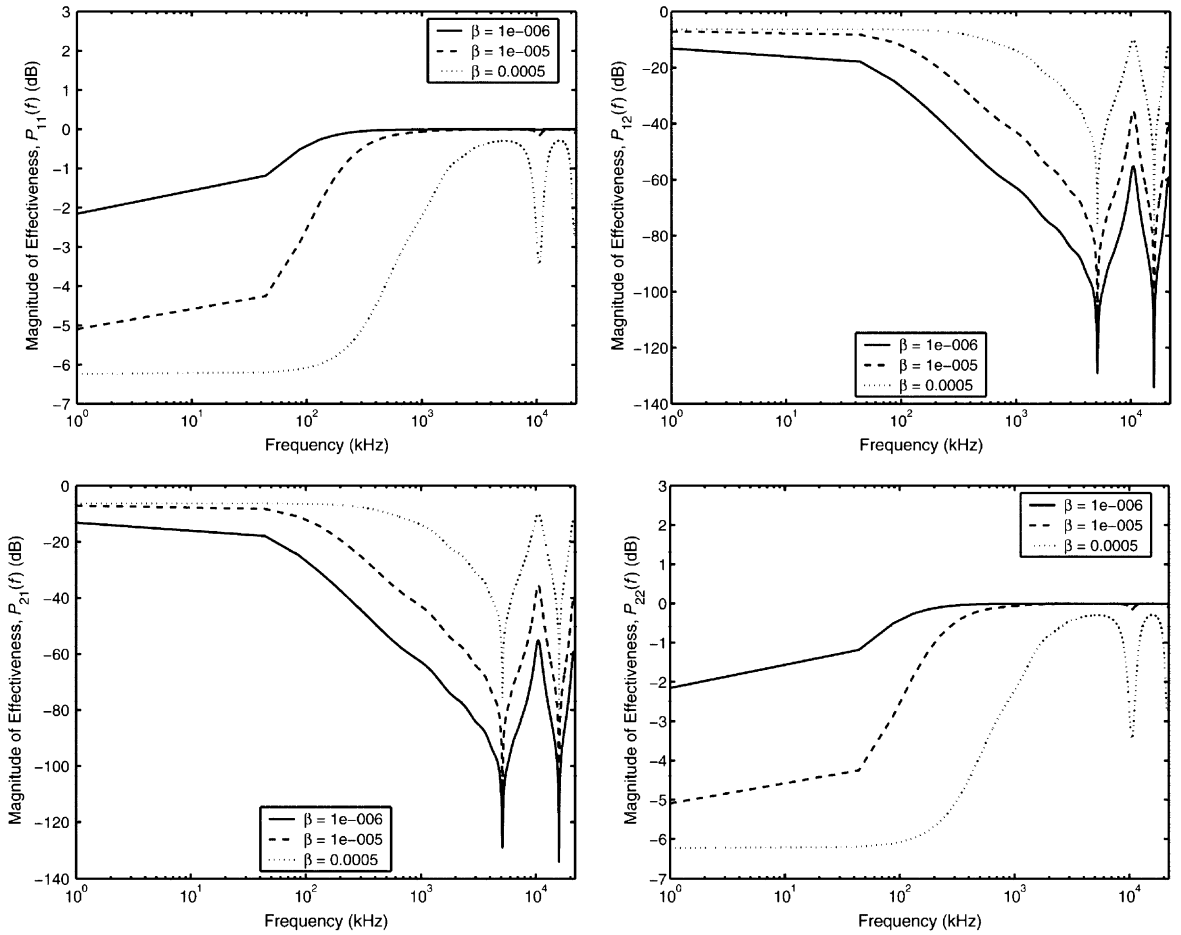

Fig. 13. The effectiveness of cross-talk cancellation of the inverse filters designed for the 10° source angle shown in the geometry of Fig. (6a). The cross-talk cancellation effectiveness are shown for values of regularisation parameter $\beta = 10^{-6}$, 10^{-5} and 0.0005.

"right ear" when the sources (loudspeakers) are spaced at a span of 60° and when the pulse is centred on the first ringing frequency of 1.65 kHz. Fig. 15 shows the sound field generated when the centre frequency of the pulse is again at 1.65 kHz but when the source span is 10°. The sound field generated is clearly more "compact" in the time domain. It has also been shown analytically in Ref. [21] that avoidance of the "ringing frequency" will generate a sound field that will be less prone to error associated with movement of the listener's head. These observations may help explain the success of the "Stereo Dipole" virtual sound imaging system and also the early observations made by Laurisden (as reported by Heegaard [38]) and observed more recently by Bauck and Cooper [10].

The general principle, that the time domain response of the reproduced field is of short duration when the inversion problem is well-conditioned can be extended to the use of other source spans. Fig. 16 shows grey-scale plots of the condition number of the matrix $\mathbf{C}(k)$ as a function of frequency and included angle between the pair of sources assuming the geometry of Fig. 6(a). Results are shown both with and without the presence of the rigid sphere. Clearly, the effect of the spherical scattering is to reduce the peaks in the condition number, but the basic dependence of condition number on loudspeaker span and frequency remains very much as predicted by the free-field model described in Ref. [25]. It was recognised by Takeuchi et al. [21–24] (and is implicit in the work of Ward and Elko [19]) that the regions in which the inversion problem becomes well-conditioned (the light areas in Figs. 16(a) and (b)) suggest that different spans of loudspeaker should be used in different frequency ranges. Clearly, the best solution to ensure a well-conditioned inversion problem is to

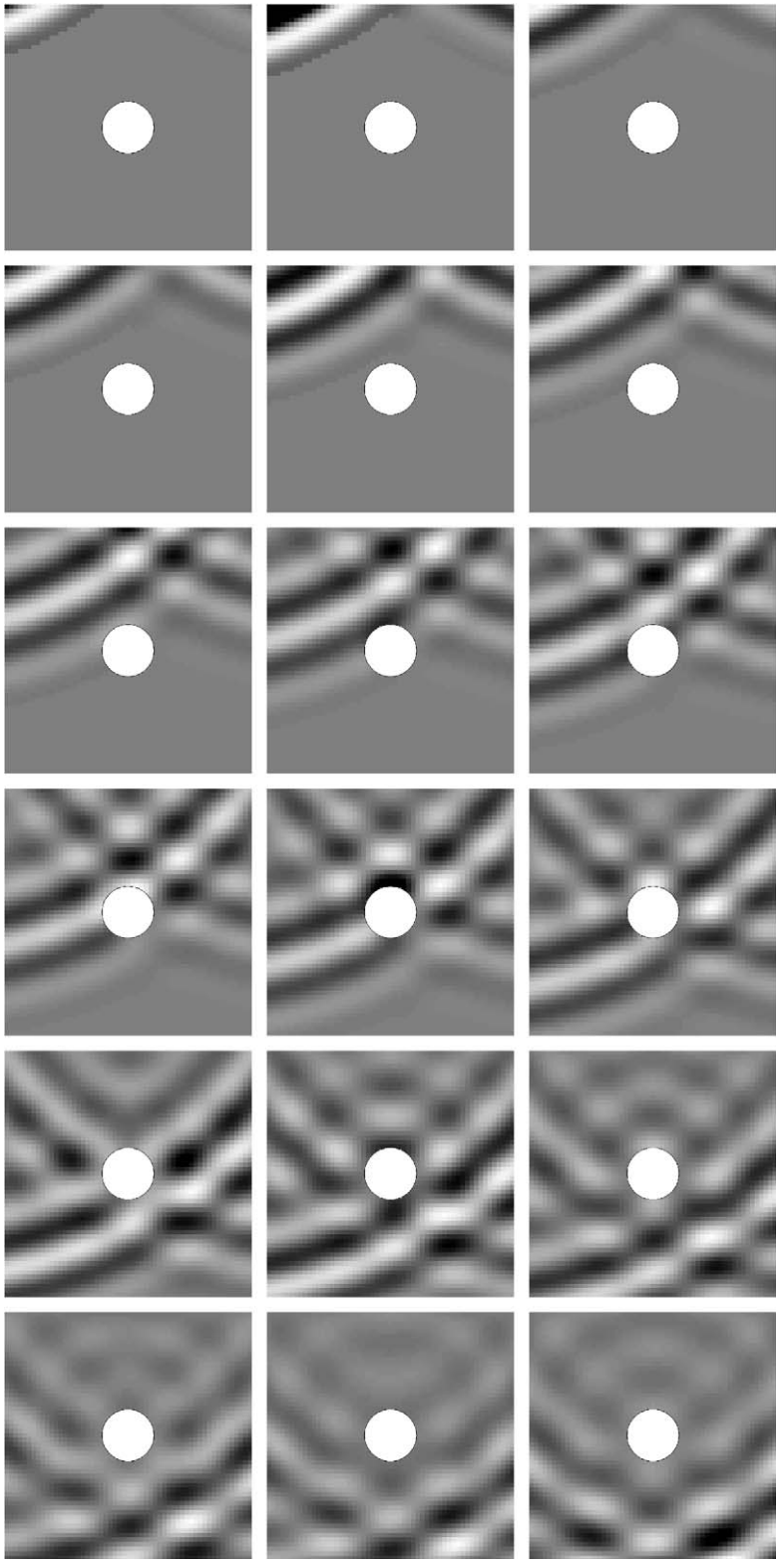

Fig. 14. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the "left ear" of the listener and a zero signal at the right ear. The geometry is that shown in Fig. 6(a) with the included angle between the sources of 60°. The centre frequency of the Gaussian pulse is 1.65 kHz, which corresponds to the first ringing frequency of the inverse filters.

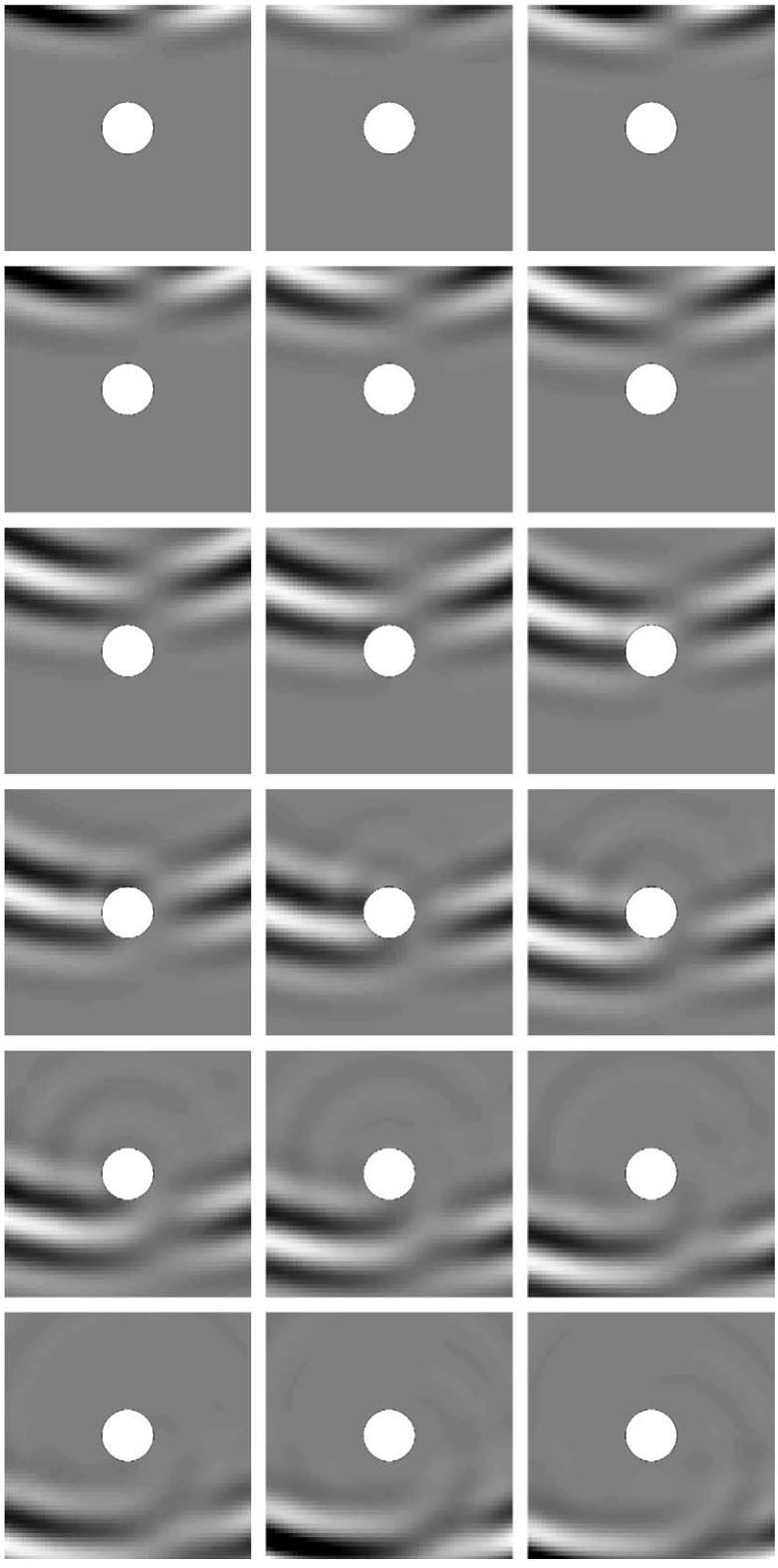

Fig. 15. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the "left ear" of the listener and a zero signal at the right ear. The geometry is that shown in Fig. 6(a) with the included angle between the sources of 10°. The centre frequency of the Gaussian pulse is 1.65 kHz, which lies within a range of frequencies for which the inversion problem is well-conditioned.

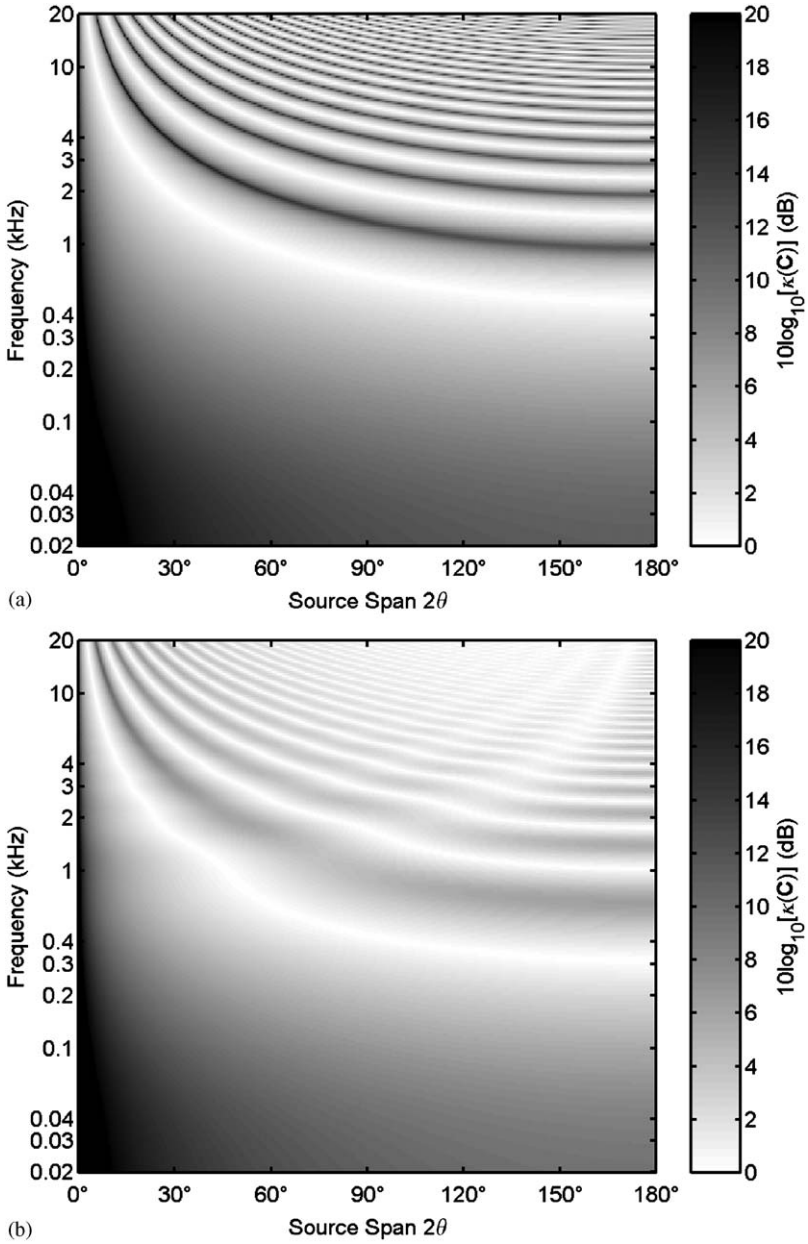

Fig. 16. A grey-scale plot of the condition number of the matrix $\mathbf{C}(k)$ when the included angle between the sources is varied in the geometry of Fig. 6(a). The results for the free-field case are shown in (a) and for the sphere scattering case in (b).

use a distribution of source strength radiating a frequency content that varies continuously with angular span. This is the so-called ''Optimal Source Distribution'' [21–24].

Figs. 17–19 show the sound field when the loudspeakers are at angles of 6°, 180° and 32°, respectively. These are the angles chosen by Takeuchi et al. [21–24] in implementing a three-loudspeaker pair realisation of the Optimal Source Distribution. In these cases the spectrum of the Gaussian pulse is chosen to be in the range of low condition number of the matrix $\mathbf{C}(k)$. Also for comparison, Fig. 20 shows the form of the sound field when the pulse is centred on the first ringing frequency associated with the 32° span. Again, the large values of condition number are associated with an undesirable response in the time domain.

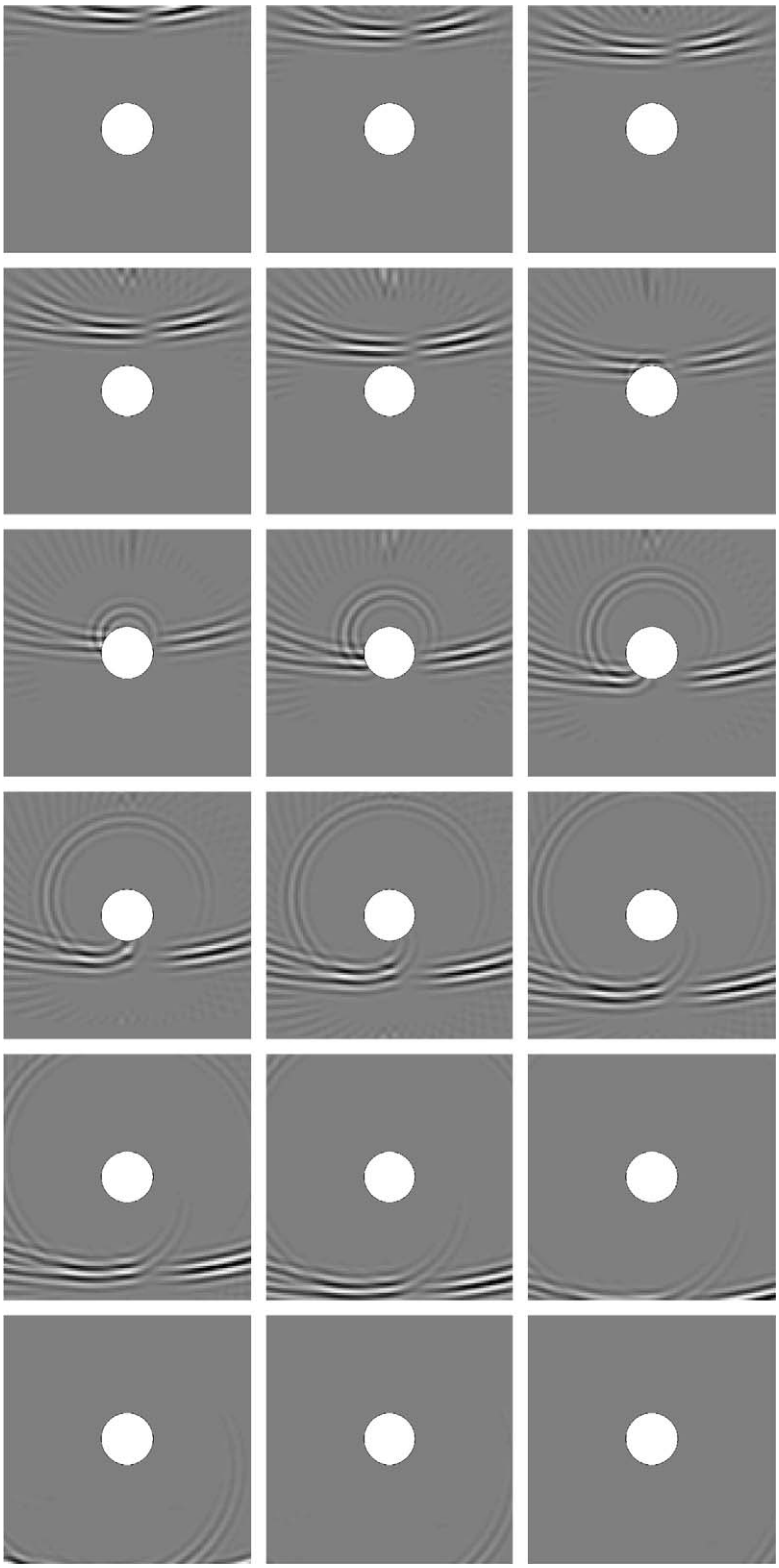

Fig. 17. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the "left ear" of the listener and a zero signal at the right ear. The geometry is that shown in Fig. 6(a) with the included angle between the sources of 6°. The centre frequency of the Gaussian pulse is 8.5 kHz, which lies within a range of frequencies for which the inversion problem is well-conditioned.

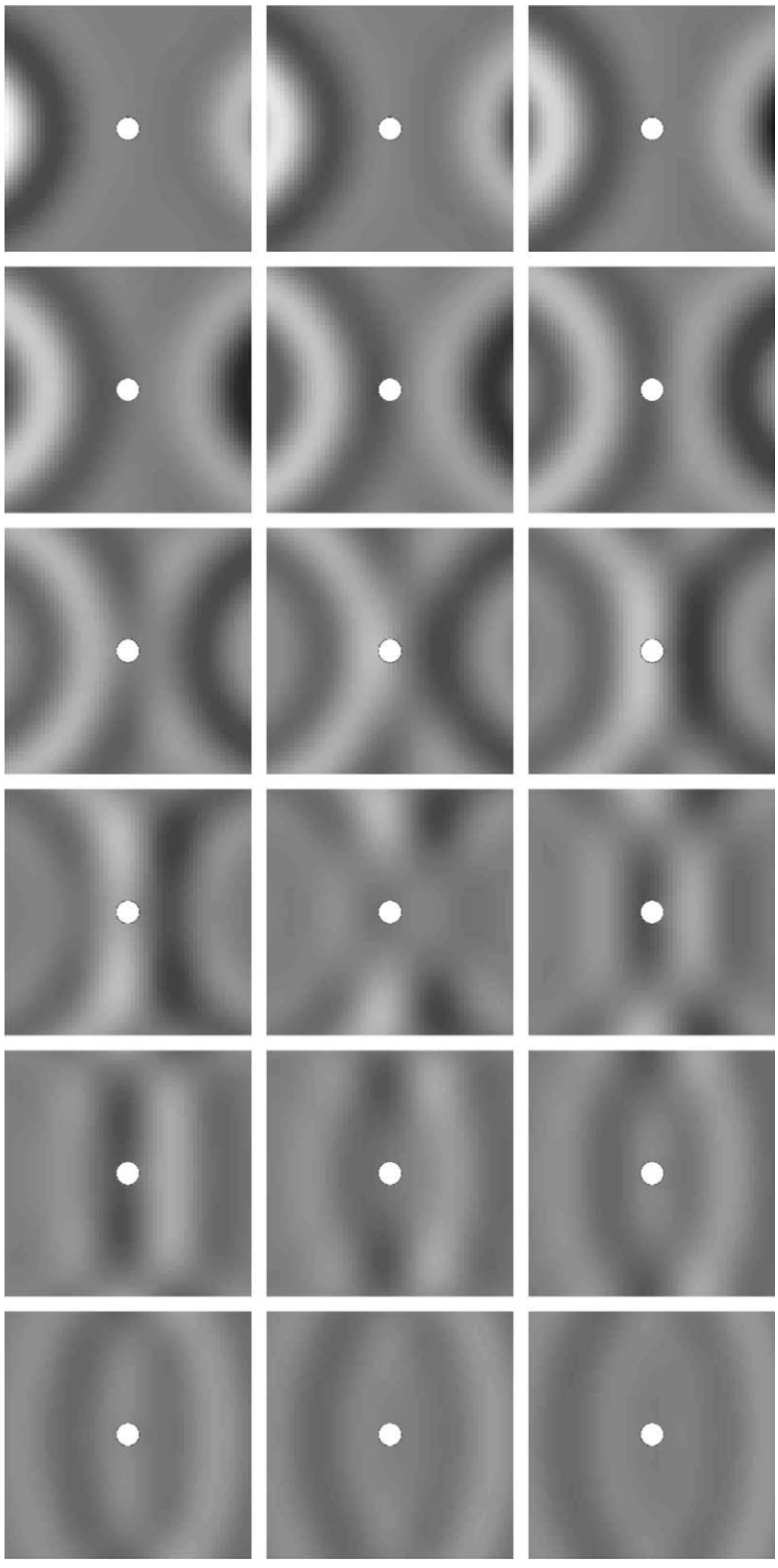

Fig. 18. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the "left ear" of the listener and a zero signal at the right ear. The geometry is that shown in Fig. 6(a) with the included angle between the sources of 180°. The centre frequency of the Gaussian pulse is 350 Hz, which lies within a range of frequencies for which the inversion problem is well-conditioned.

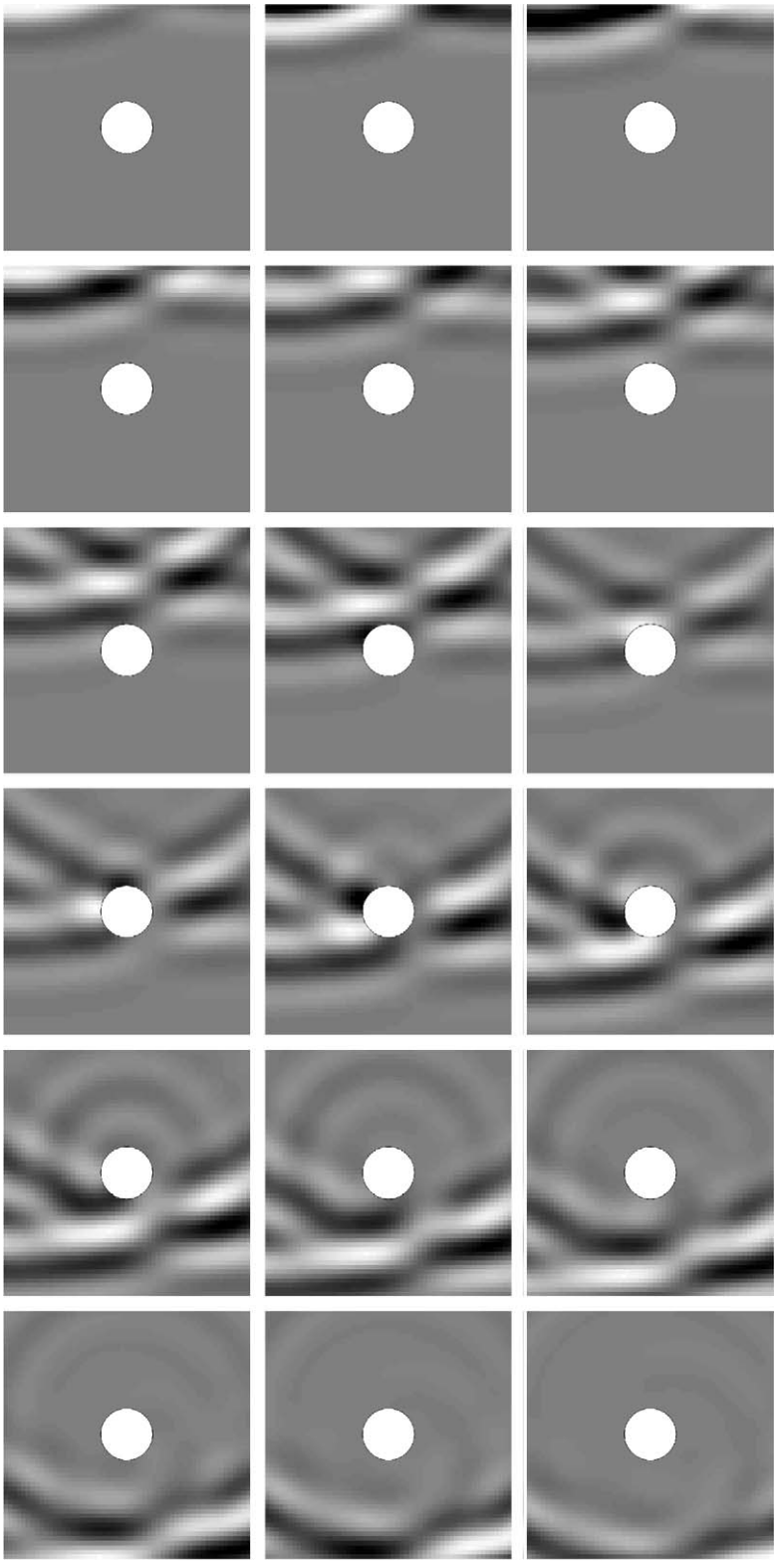

Fig. 19. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the "left ear" of the listener and a zero signal at the right ear. The geometry is that shown in Fig. 6(a) with the included angle between the sources of 32°. The centre frequency of the Gaussian pulse is 1.5 kHz, which lies within a range of frequencies for which the inversion problem is well-conditioned.

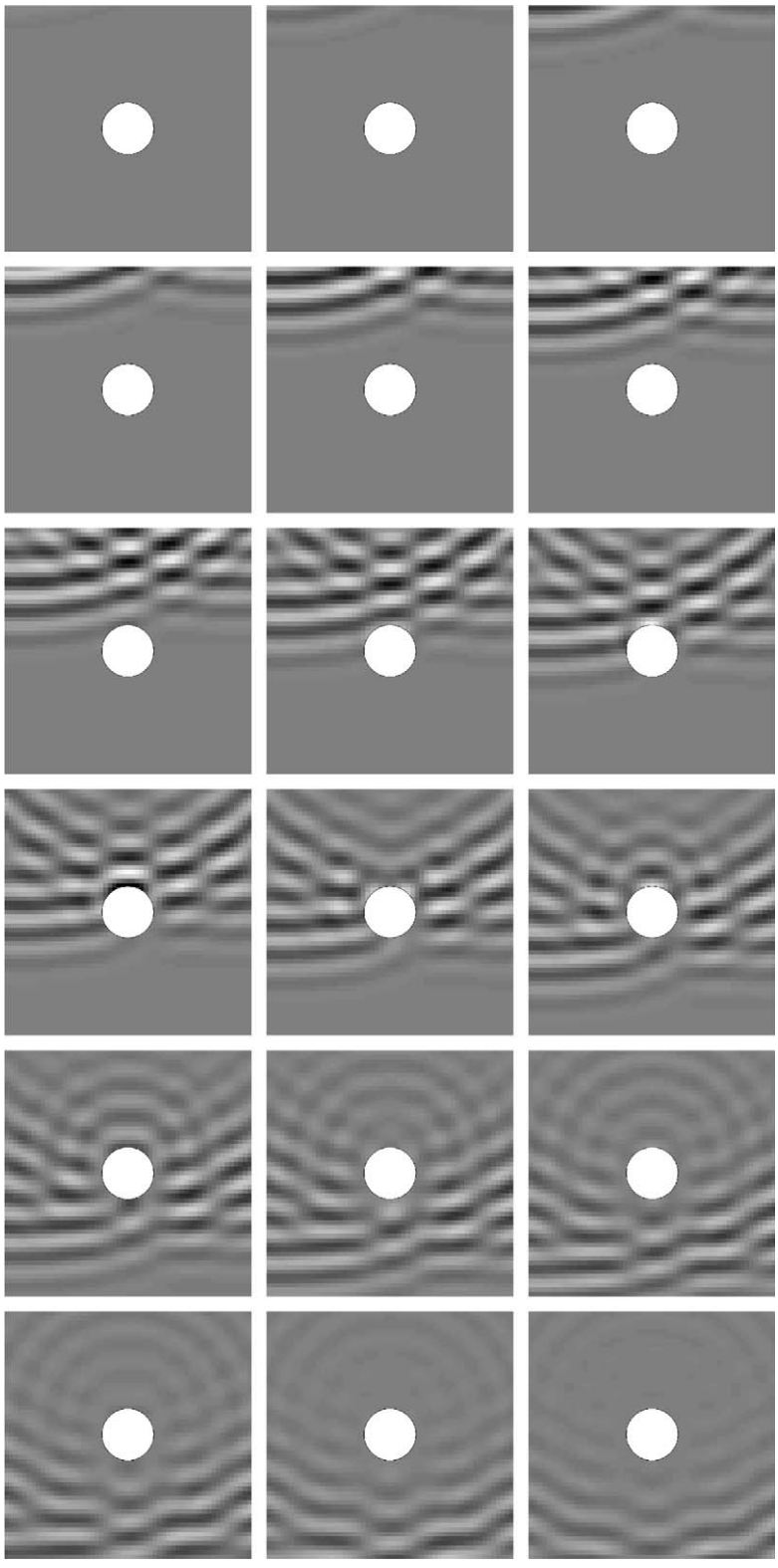

Fig. 20. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the "left ear" of the listener and a zero signal at the right ear. The geometry is that shown in Fig. 6(a) with the included angle between the sources of 32°. The centre frequency of the Gaussian pulse is 3.11kHz, which lies within a range of frequencies for which the inversion problem is *not* well-conditioned.

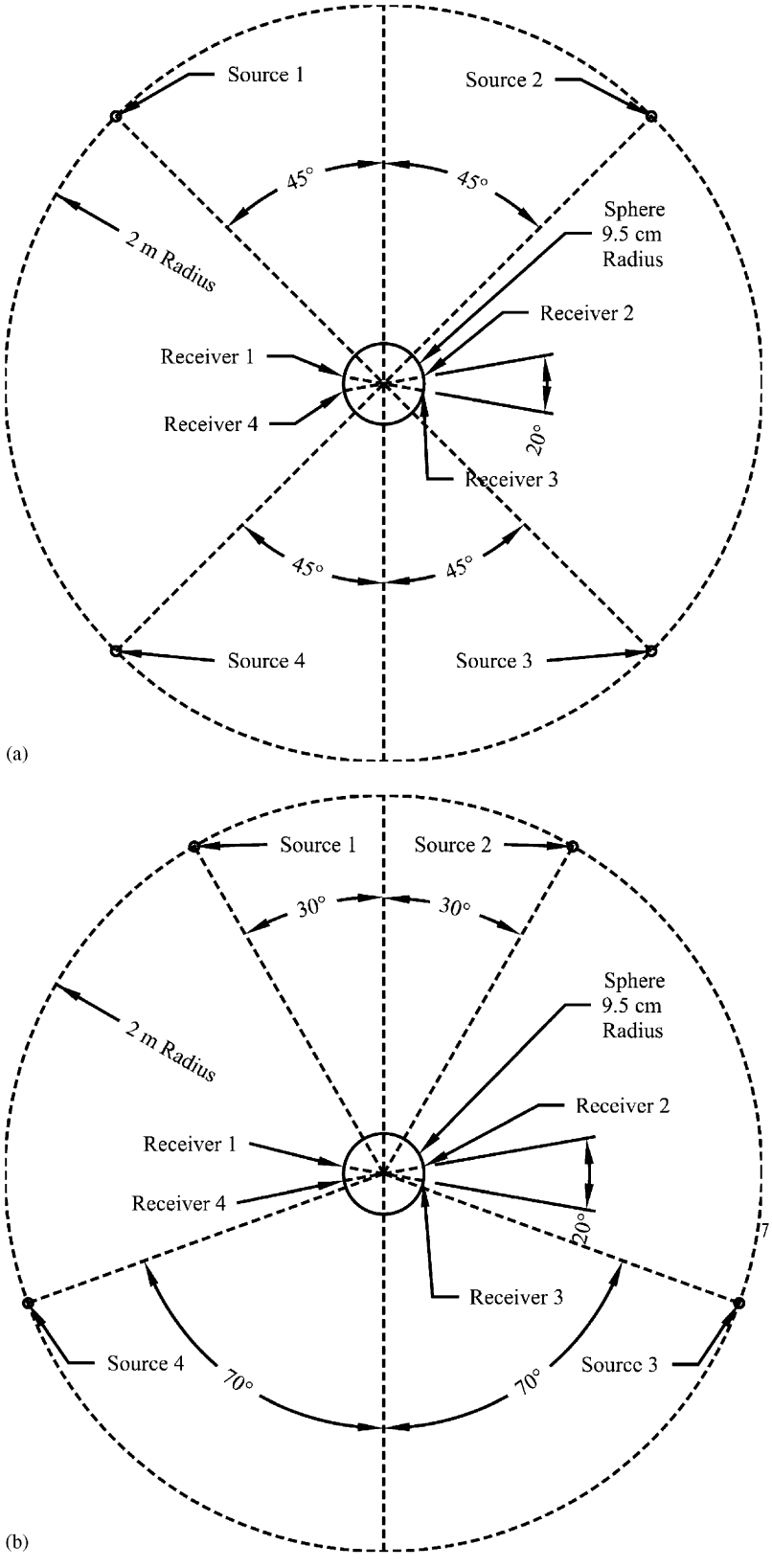

Fig. 21. The arrangement of loudspeakers and microphones used in the virtual acoustic imaging system investigated by Kahana et al. [39] showing (a) the symmetric and (b) the asymmetric loudspeaker arrangements.

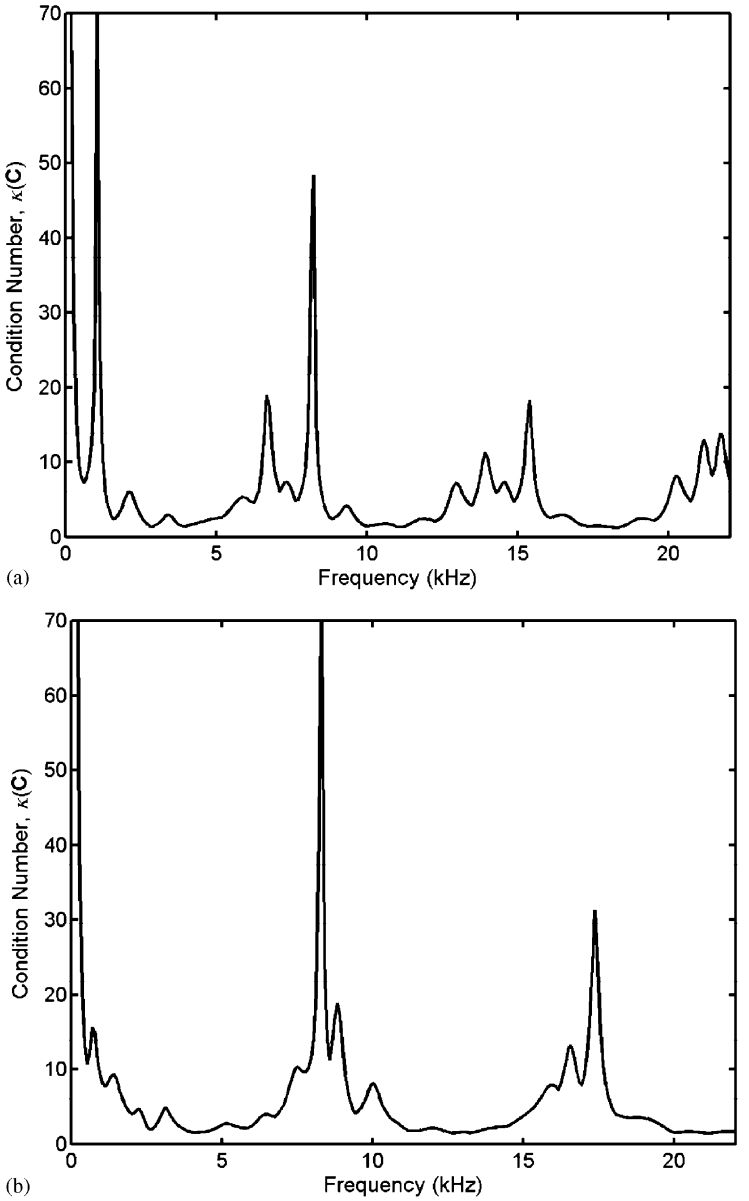

Fig. 22. Condition numbers of the matrix $\mathbf{C}(k)$ associated with the geometry of Fig. 21 when the loudspeaker spans used are (a) those shown in Figs. 21(a) and (b) those shown in Fig. 21(b).

8. The time domain response of multi-channel systems

Another virtual imaging system that has been found to operate well is that described by Kahana et al. [39]. The general layout of the loudspeakers relative to the listener is illustrated in Fig. 21. The system made use of an arrangement of two loudspeakers to the front of the listener and two to the rear. A matrix of cross-talk cancellation filters was designed that operated on the four signals recorded by two pairs of microphones placed on either side of a rigid sphere, the microphones in each pair being relatively close together compared to the acoustic wavelength in the range of frequencies of interest. The cross-talk cancellation matrix was designed to ensure the reproduction of these recorded signals in the region of the ears of the listener. The rationale behind this design was an attempt to solve the difficulty of "front–back confusion" experienced with two channel systems. This occurs when small rotations of the listener's head result in virtual images intended for the rear of

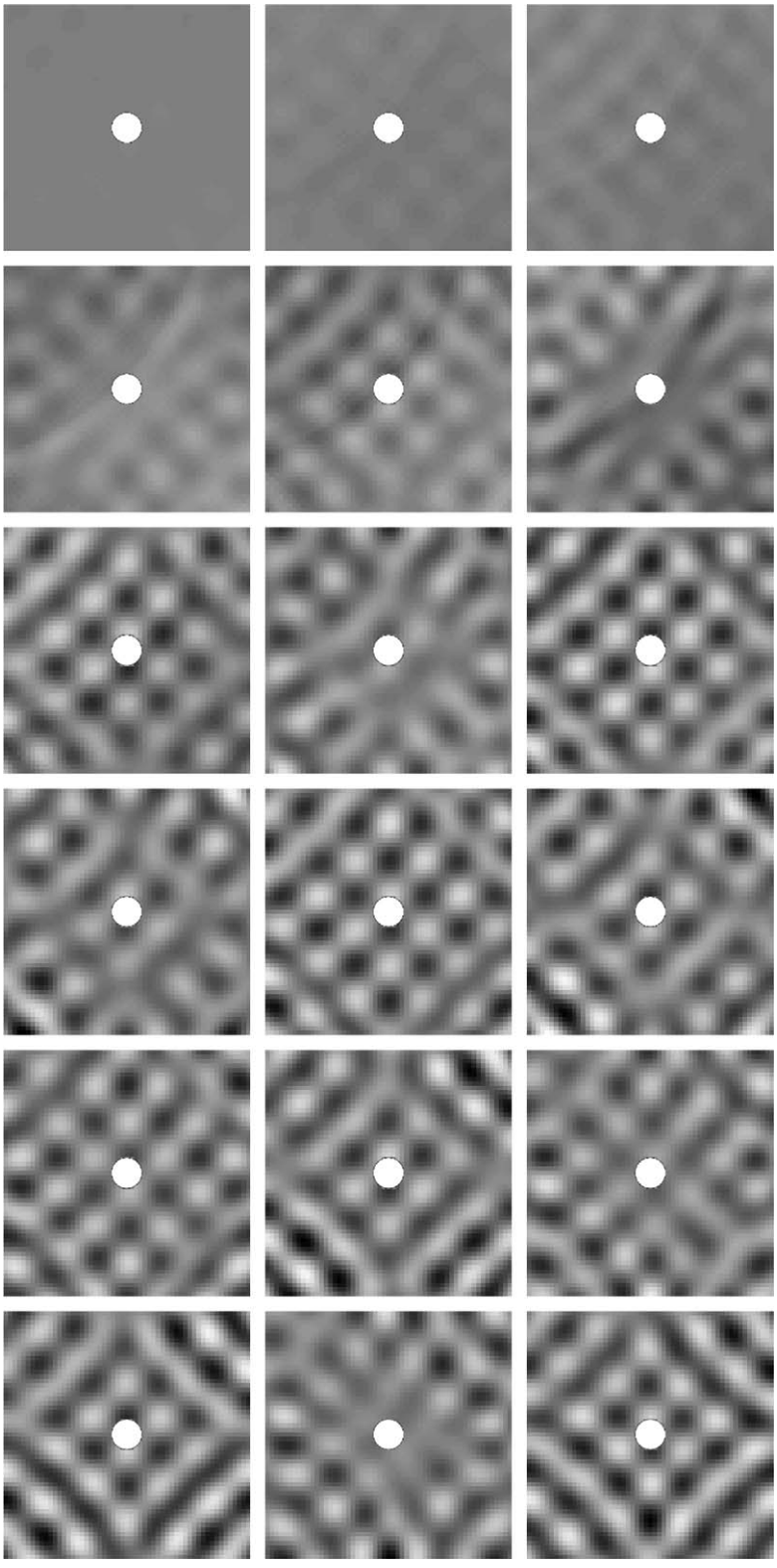

Fig. 23. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the front microphone on the left side of the listener and a zero signal at the other three microphones. The geometry is that shown in Fig. 21 with an included angle between the front and rear sources of 90°. The centre frequency of the Gaussian pulse is 1.04 kHz, which lies within a range of frequencies for which the inversion problem is *not* well-conditioned.

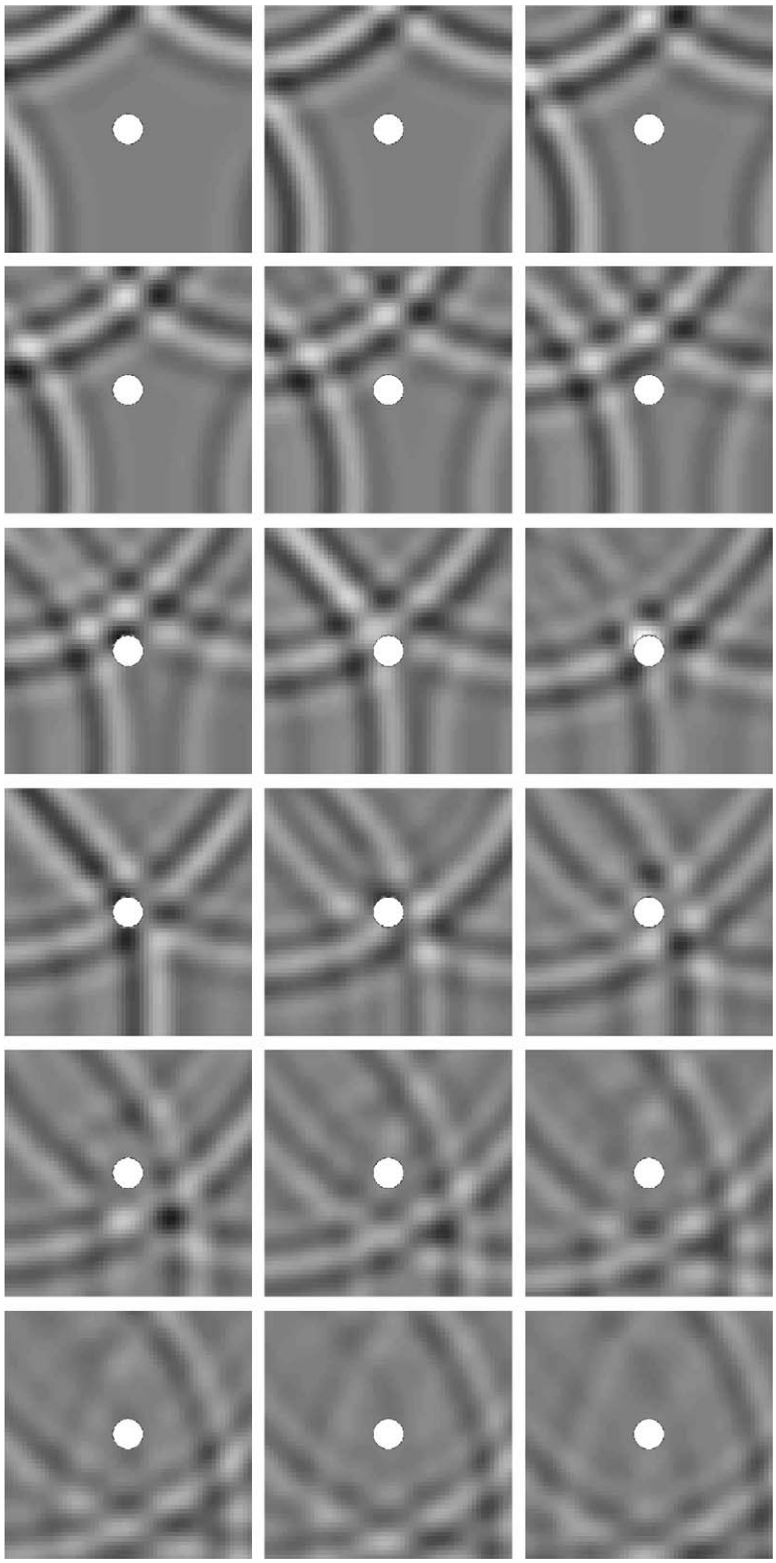

Fig. 24. The evolution of the sound field when the desired signals consist of a Gaussian pulse at the front microphone on the left side of the listener and a zero signal at the other three microphones. The geometry is that shown in Fig. 21 with an included angle between the front sources of 60° and an included angle between the rear sources of 140°. The centre frequency of the Gaussian pulse is 1.04 kHz, which lies within a range of frequencies for which the inversion problem is well-conditioned.

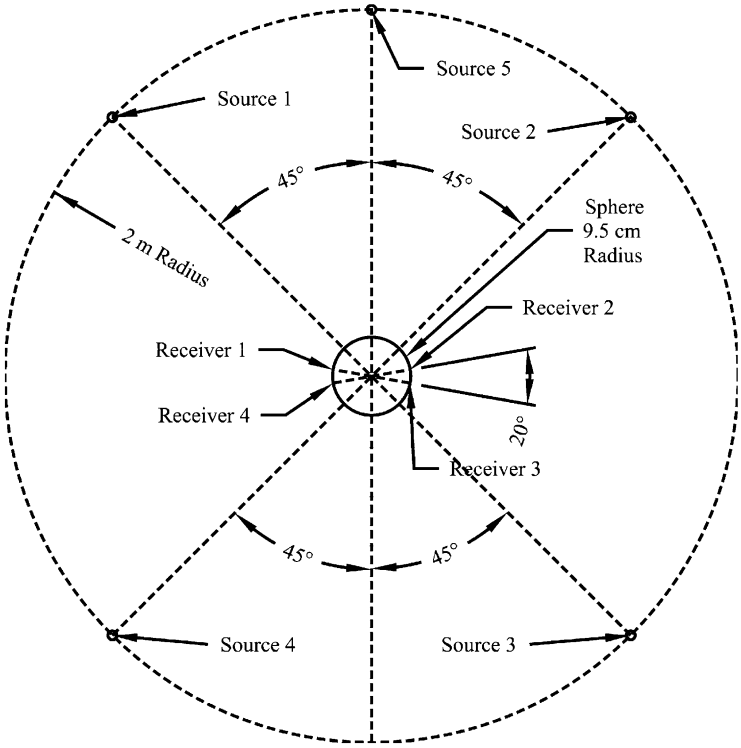

Fig. 25. A five-source loudspeaker arrangement for the reproduction of recorded signals at four microphones.

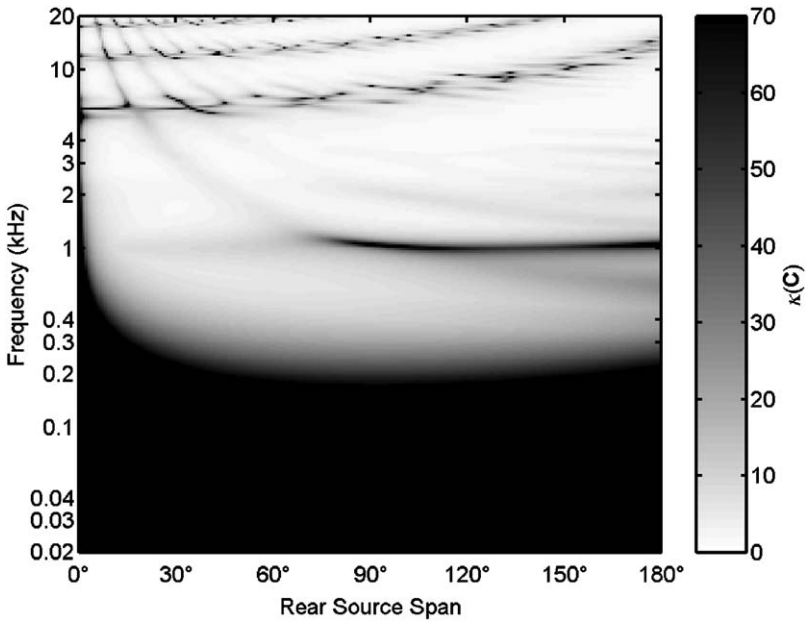

Fig. 26. Condition number of the matrix $\mathbf{C}(k)$ for the four-source arrangement with a 90° front source span while the rear source span varies between 0° and 180°.

the listener being perceived to the front of the listener at an equivalent "mirrored" angular location. The motivation for the system design was the replication at the listeners ears of not only the pressure, but also the pressure gradient in the sound field, on the assumption that correct replication of this quantity would enable

small movements of the listener's head whilst still preserving a robust and stable image. The system implemented by Kahana et al. [39] was found to resolve the front–back ambiguity for the speech signals recorded in practice, and the basic properties of such a system were subsequently confirmed in subjective experiments reported by Hill et al. [40]. However, the full characteristics of this system were not thoroughly explored and it is interesting to re-examine these here.

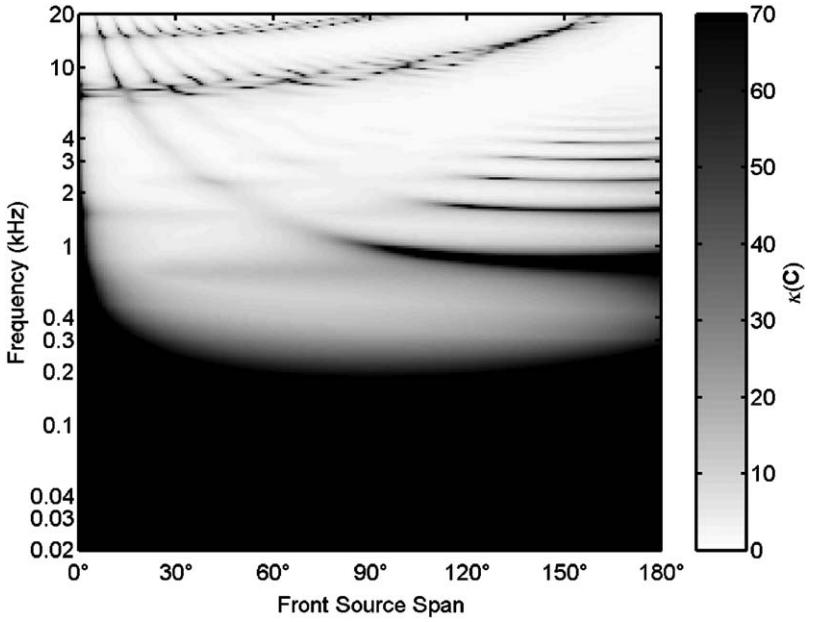

Fig. 27. Condition number of the matrix $\mathbf{C}(k)$ for four-source arrangement with a 140° rear source span while the front source span varies between 0° and 180°.

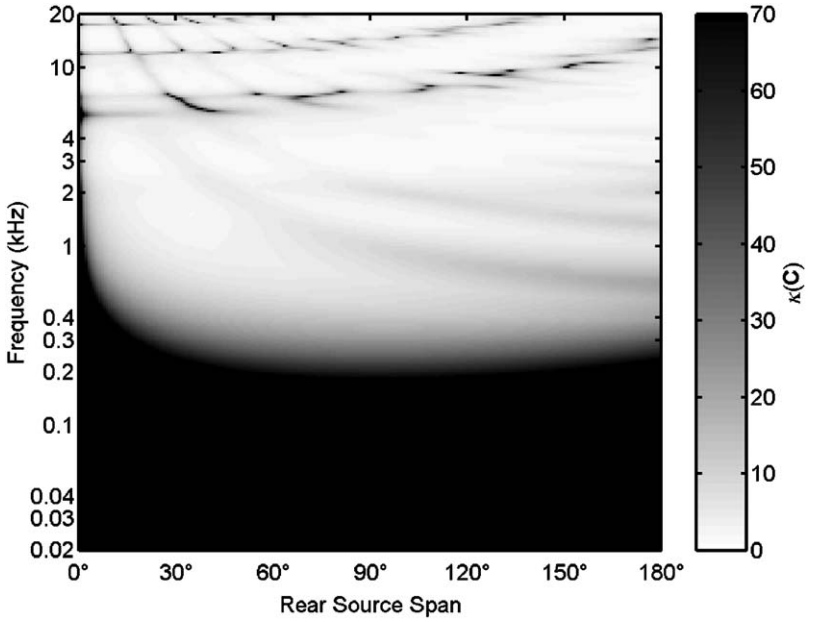

Fig. 28. Condition number of the matrix $\mathbf{C}(k)$ for the five-source arrangement with a 90° front source span while the rear source span varies between 0° and 180°.

The condition number of the matrix $\mathbf{C}(k)$ of transfer functions appropriate to the geometry of Fig. 21 is plotted in Fig. 22 for two choices of loudspeaker layout. The first "symmetric" loudspeaker layout made use of loudspeaker spans of 90° at both front and rear whilst a second "asymmetric" arrangement used a span of 60° to the front and a span of 140° to the rear. Fig. 22 shows that the condition number of the symmetric arrangement becomes particularly large in the region of 1 kHz whilst the asymmetric arrangement produces a low condition number over a reasonable bandwidth until a significant peak occurs in the condition number at about 8 kHz. Figs. 23 and 24 illustrate the form of the sound field generated when cross-talk cancellation is attempted with a Gaussian pulse centred on 1.04 kHz for each of the arrangements of sources. The time domain response of the symmetric arrangement is far less well-contained than that of the asymmetric arrangement, the ringing in the sound field being clearly evident at this frequency. It is interesting to note that Kahana et al. [39] reported that the symmetric arrangement of loudspeakers did not produce good results in informal listening tests, with images being "localised inside the head". The main subjective experiments were therefore undertaken with the asymmetric loudspeaker arrangement and this was found to perform well in resolving front–back ambiguity.

With the increasing prevalence of five loudspeaker systems, it is also interesting to explore the characteristics of potential systems that make use of this standard layout together with a specific recording system aimed at accurate signal reproduction in the region of the listener's head. One approach is to use the essence of the approach defined by Kahana et al. [39] but modify the system to make use of an additional loudspeaker to the front of the listener. One can then use the solution for the cross-talk cancellation matrix that results from ensuring zero error at the four positions in the reproduced field but whilst using minimum sum of squared loudspeaker input voltages. The solution of this constrained optimisation problem [34] results in the cross-talk cancellation matrix given by

$$\mathbf{H}_x(k) = \mathbf{C}^{\mathrm{H}}(k)\left[\mathbf{C}(k)\mathbf{C}^{\mathrm{H}}(k)\right]^{-1} \mathrm{e}^{-\mathrm{j}\omega\Delta}. \tag{32}$$

It is first interesting to observe the influence on the condition number of the matrix $\mathbf{C}(k)$ of the introduction of an additional loudspeaker. The geometry of Fig. 25 has been assumed, with the introduction of a fifth loudspeaker on the same arc as the other loudspeakers, but placed directly to the front of the listener. Fig. 26 shows the condition number of the matrix $\mathbf{C}(k)$ for the four source arrangement with the front source span fixed at 90° with the rear source span varied between 0°and 180° (Fig. 27). Fig. 28 shows the same plot but

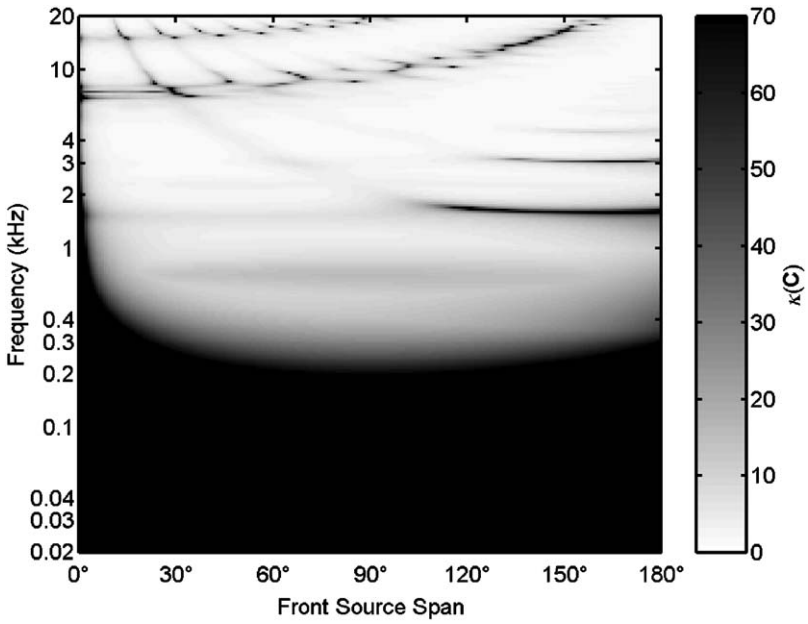

Fig. 29. Condition number of the matrix $\mathbf{C}(k)$ for the five-source arrangement with a 140° rear source span while the front source span varies between 0° and 180°.

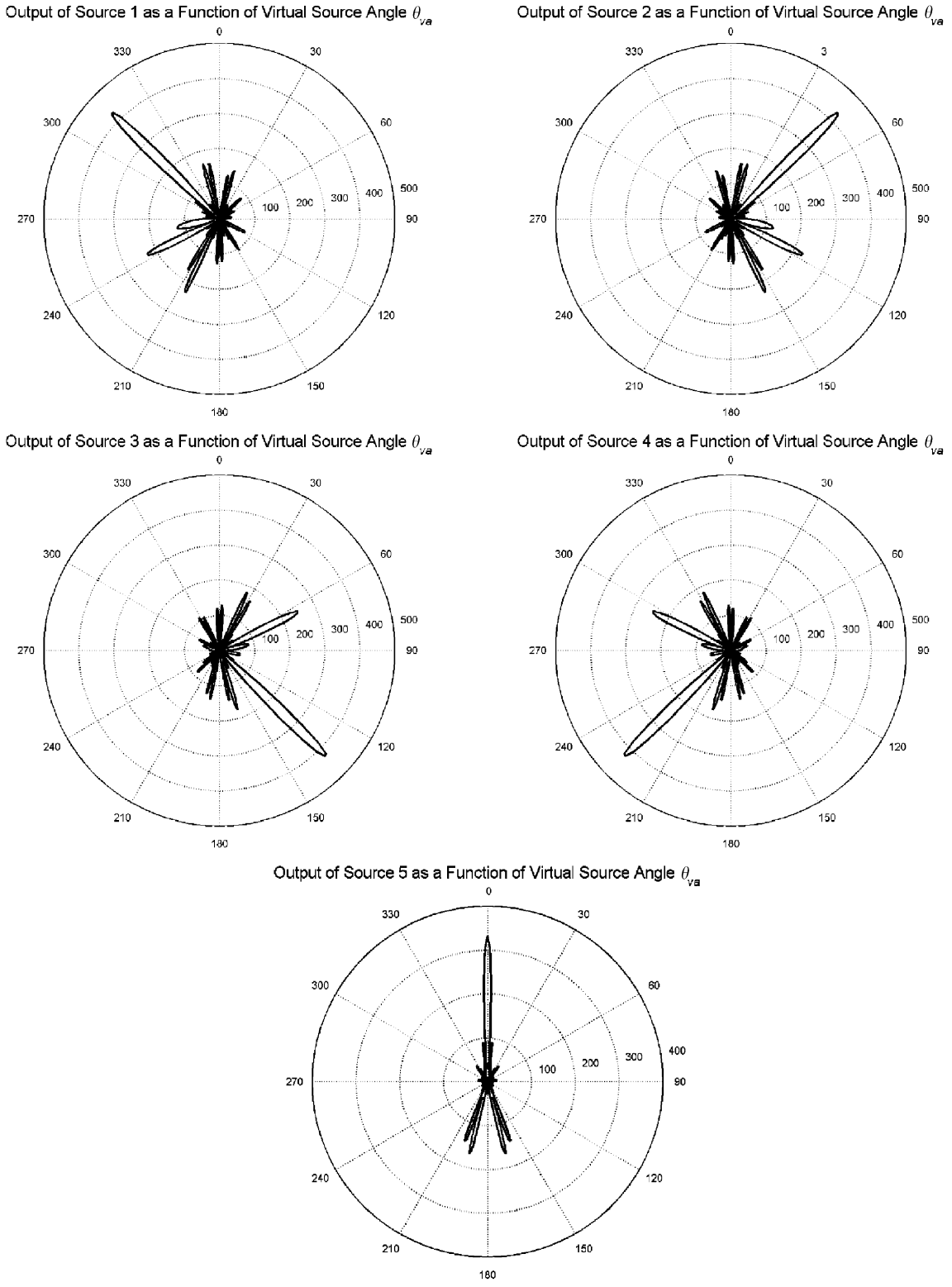

Fig. 30. The outputs of the five sources in the geometrical arrangement of Fig. 25 as a function of the angular position of a source moved relative to the sphere upon which the acoustic pressure is recorded in order to define the desired signals. Note that the outputs of the respective sources are a maximum when the angle of the source of the recorded signals coincides with the angle of the sources used for reproduction.

with the addition of the fifth loudspeaker. Clearly several regions of high condition number have been reduced by the addition of the further loudspeaker. Similarly, Fig. 27 shows the condition number of $\mathbf{C}(k)$ as the rear source span is fixed at 140° whilst the front source span is varied between 0° and 180°, and Fig. 29 shows that the addition of the further loudspeaker reduces regions of high condition number. It therefore appears from these considerations that the system developed by Kahana et al. [39] might perhaps be improved upon by the addition of a further source to the front of the listener. Fig. 30 shows the outputs of the five sources as a function of the angular position of a source used to generate the four signals recorded on the surface of the rigid sphere. These outputs were computed following the calculation of the cross-talk cancellation matrix in accordance with Eq. (32). It is evident from these results that when sound radiated from a particular angular direction is to be reproduced, then the source closest to that direction produces the dominant output. One might expect that this is a desirable feature of a multi-channel recording and reproduction system. Note also that there is left–right symmetry in the results that reflects the symmetry in the loudspeaker arrangement, but that there is front–back asymmetry in the results that reflects the front–back asymmetry in the loudspeaker arrangement. These observations may prove useful in the further development of multi-channel recording techniques suitable for five channel reproduction systems.

9. Conclusions

It has been demonstrated that there is a strong link between the time domain response of a number of systems for generating virtual acoustic images and the condition number of the matrix of transfer functions that must be inverted in order to achieve cross-talk cancellation. Although a previous paper [25] demonstrated this relationship analytically in the free-field two-source/two-field point case, it has been found to hold in a number of other cases where the influence of the listener's head is modelled as a rigid sphere. It has been assumed that the time domain response produced at the "ringing frequencies" of the reproduction systems is psychoacoustically undesirable, although there is not yet incontrovertible experimental evidence to suggest that this is necessarily the case. It does seem, however, that subjectively successful virtual acoustic imaging systems exhibit a time domain response of short duration in the frequency bands of successful operation.

Acknowledgements

The authors wish to express their sincere gratitude to Mr. Timos Papadopoulos who undertook the numerical simulations leading to the results presented in Fig. 7.

References

[1] A.D. Blumlein, Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems, British Patent No. 394,325, 1931.
[2] B.S. Atal, M.R. Schroeder, Apparent sound source translator, US Patent No. 3,236,949, 1966.
[3] B.B. Bauer, Stereophonic earphones and binaural loudspeakers, *Journal of the Audio Engineering Society* 9 (1961) 148–151.
[4] M.R. Schroeder, D. Gottlob, K.F. Siebrasse, Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters, *Journal of the Acoustical Society of America* 56 (1974) 1195–1201.
[5] P. Damaske, V. Mellert, Sound reproduction of the upper semi-space with directional fidelity using two loudspeakers, *Acustica* 22 (1969) 153–162 (In German).
[6] H. Hamada, N. Ikeshoji, Y. Ogura, T. Miura, Relation between physical characteristics of orthostereophonic system and horizontal plane localization, *Journal of the Acoustical Society of Japan* 6 (1985) 143–154.
[7] G. Neu, E. Mommertz, A. Schmitz, Investigations on true directional sound reproduction by playing head-referred recordings over two loudspeakers: Part 1, *Acustica* 76 (1992) 183–192 (In German).
[8] G. Urbach, E. Mommertz, A. Schmitz, Investigations on the directional scattering of sound reflections from the playback of head-referred recordings over two loudspeakers: Part 2, *Acustica* 77 (1992) 153–161 (In German).
[9] D.H. Cooper, J.L. Bauck, Prospects for transaural recording, *Journal of the Audio Engineering Society* 37 (1989) 3–19.
[10] D.H. Cooper, J.L. Bauck, Generalised transaural stereo and applications, *Journal of the Audio Engineering Society* 44 (1992) 683–705.
[11] D.H. Cooper, J.L. Bauck, Head diffraction compensated stereo system with loudspeaker array, US Patent No. 5,333,200 1994.
[12] P.A. Nelson, H. Hamada, S.J. Elliott, Adaptive inverse filters for stereophonic sound reproduction, *Institute of Electrical and Electronics Engineers Transactions on Acoustics, Speech and Signal Processing* 40 (1992) 1621–1632.

[13] P.A. Nelson, F. Orduna-Bustamante, D. Engler, Multi-channel signal processing techniques in the reproduction of sound, *Journal of the Audio Engineering Society* 44 (1996) 973–989.
[14] W.G. Gardner, 3-D Audio Using Loudspeakers, PhD Thesis, MIT Media Laboratory, Cambridge, MA, 1997.
[15] P.A. Nelson, O. Kirkeby, T. Takeuchi, H. Hamada, Sound fields for the production of virtual acoustic images. Letter to the Editor, *Journal of Sound and Vibration* 204 (2) (1997) 386–396.
[16] O. Kirkeby, P.A. Nelson, H. Hamada, The stereo dipole—a virtual source imaging system using two closely spaced loudspeakers, *Journal of the Audio Engineering Society* 46 (5) (1998) 387–395.
[17] O. Kirkeby, P.A. Nelson, Local sound field reproduction using two closely spaced loudspeakers, *Journal of the Acoustical Society of America* 104 (3) (1998) 1973–1981.
[18] T. Takeuchi, P.A. Nelson, Robustness to head misalignment of virtual sound imaging systems, *Journal of the Acoustical Society of America* 109 (2001) 958–971.
[19] D.B. Ward, G.W. Elko, Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation, *Institute of Electrical and Electronics Engineers Signal Processing Letters* 6 (5) (1999) 106–108.
[20] J.L. Bauck, A simple loudspeaker array and associated crosstalk canceller for improved 3D audio, *Journal of the Audio Engineering Society* 49 (2001) 3–13.
[21] T. Takeuchi, P.A. Nelson, Optimal source distribution for virtual acoustic imaging, Institute of Sound and Vibration Research Technical Report No. 288, University of Southampton, 2000.
[22] T. Takeuchi, M. Teschl, P.A. Nelson, Subjective evaluation of the optimal source distribution system for virtual acoustic imaging, *Proceedings of the 19th Audio Engineering Society International Conference*, Schloss Elmau, Germany, June 2001.
[23] T. Takeuchi, P.A. Nelson, Optimal source distribution for virtual acoustic imaging, *Proceedings of the Audio Engineering Society 110th Convention,* Amsterdam, May 2001, preprint No. 5372.
[24] T. Takeuchi, Systems for Virtual Acoustic Imaging Using the Binaural Principle, PhD Thesis, University of Southampton, 2001.
[25] P.A. Nelson, Active control for virtual acoustics, *Proceedings of Active 2002*: *The 2002 International Symposium on Active Control of Sound and Vibration* **1**, July 2002, pp. 67–89.
[26] Lord Rayleigh, *The Theory of Sound*, second rev. ed., Vol. 2, Dover Publications, New York, 1945.
[27] Lord Rayleigh (J.W. Strutt), On our perception of sound direction, *Philosophical Magazine* 13 (1907) 214–232.
[28] E.R. Hafter, C. Trahiotis, Functions of the binaural system, in: M.J. Crocker (Ed.), *Encyclopaedia of Acoustic*, Vol. 3, Wiley, New York, 1997, pp. 1461–1479.
[29] D.W. Grantham, Spatial hearing and related phenomena, in: B.J.C. Moore (Ed.), *Hearing*, Academic Press, London, 1995, pp. 297–345 (Chapter 9).
[30] F.L. Wightman, D.J. Kistler, The dominant role of low-frequency interaural time differences in sound localisation, *Journal of the Acoustical Society of America* 91 (1992) 1648–1661.
[31] M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1965.
[32] J.F.W. Rose, Visually Adaptive Virtual Acoustic Imaging, PhD Thesis, University of Southampton, England, 2004.
[33] G.H. Golub, C.F. Van Loan, *Matrix Computations*, third ed., The Johns Hopkins University Press, Baltimore, MD, 1996.
[34] P.A. Nelson, S.J. Elliott, *Active Control of Sound*, Academic Press, London, 1992.
[35] O. Kirkeby, P.A. Nelson, H. Hamada, F. Orduna-Bustamante, Fast deconvolution of multichannel systems using regularization, *Institute of Electrical and Electronics Engineers Transactions on Speech and Audio Processing* 6 (2) (1998) 189–194.
[36] J.G. Proakis, D.G. Manolakis, *Digital Signal Processing Principles, Algorithms and Applications*, second ed., Macmillan, New York, 1992.
[37] D. Gabor, Theory of Communication, *Journal of the Institute of Electrical and Electronics Engineers*, London, 93(III) (1946), 429–457.
[38] Fr.D. Heegard, *EBU Rev. pt A—Technical*, No.52 (1958), 2-6. (reprinted in *Journal of the Audio Engineering Society* 40 (1992) 692–705).
[39] Y. Kahana, P.A. Nelson, O. Kirkeby, H. Hamada, A multiple microphone recording technique for the generation of virtual acoustic images, *Journal of the Acoustical Society of America* 105 (1999) 1503–1516.
[40] P.A. Hill, P.A. Nelson, O. Kirkeby, H. Hamada, Resolution of front-back confusion in virtual acoustic imaging systems, *Journal of the Acoustical Society of America* 108 (2001) 2901–2910.